



THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par : *l'Institut National Polytechnique de Toulouse (INP Toulouse)*

Présentée et soutenue le *08/10/2021* par :

Mme YOUSRA HAMROUNI BERKAOUI

**Développement d'une approche opérationnelle pour l'identification
automatique des peupleraies à large échelle par
télédétection hypertemporelle**

De l'adaptation de domaine à la création d'un indice spectral dédié

JURY

| | | |
|------------------|--|---------------|
| PHILIPPE LEJEUNE | Professeur Gembloux Agro-Bio Tech - ULiège | Rapporteur |
| AGNÈS BÉGUÉ | Directrice de Recherche CIRAD | Rapporteuse |
| THOMAS CORPETTI | Directeur de Recherche CNRS | Examinateur |
| XAVIER BRIOTTET | Directeur de Recherche ONERA | Examinateur |
| DAVID SHEEREN | Maître de conférences INP Toulouse | Encadrant |
| VÉRONIQUE CHÉRET | Enseignante Chercheuse EI Purpan | Co-directrice |
| CLAUDE MONTEIL | Maître de conférences INP Toulouse | Co-directeur |
| ERIC PAILLASSA | Ingénieur CNPF-IDF | Membre invité |

École doctorale :

Sciences Écologiques, Vétérinaires, Agronomiques et Bioingénieries (SEVAB)

Discipline ou spécialité :

Agrosystèmes, écosystèmes et environnement

Unité de Recherche :

Dynamique et Écologie des Paysages Agriforestiers (DYNAFOR)

Résumé

Le peuplier cultivé constitue la première essence de feuillus plantée en France. Il joue un rôle économique de premier plan, notamment pour la production de bois-matériau, contreplaqués et emballages légers. Il est également valorisé dans l'industrie papetière. Pourtant, malgré l'importance de la filière, les surfaces couvertes en peupleraies à l'échelle nationale sont encore très incertaines. Selon la source de données utilisée (BD Forêt IGN, cadastre, inventaire forestier), les estimations montrent des écarts de plus de 50 000 ha. La fréquence de mise à jour de ces sources est inadaptée pour suivre le peuplier cultivé dont le cycle de rotation est court (15-20 ans).

La télédétection satellitaire est utilisée depuis longtemps pour cartographier les milieux forestiers, qu'il s'agisse de forêts naturelles ou de plantations. Avec l'amélioration constante des caractéristiques spatiale, spectrale et temporelle des capteurs, il est possible d'envisager son appropriation dans un contexte opérationnel, pour un suivi régulier de la ressource sur de grandes étendues.

L'objectif de cette thèse est double. Le premier est d'explorer le potentiel des séries temporelles d'images optiques Sentinel-2 pour distinguer automatiquement les peupleraies des autres essences de feuillus en tenant compte de la diversité des contextes populicoles. Le second est de proposer une stratégie de classification à l'échelle nationale en tenant compte de la non-stationnarité spatiale de la réponse spectrale des peupleraies, de l'hétérogénéité des acquisitions, et du nombre limité de données de référence.

La démarche adoptée a consisté à étudier différentes techniques d'adaptation de domaine disponibles dans le champ de l'apprentissage automatique. Ces techniques, non supervisées ou semi-supervisées, ont permis de répondre aux contraintes de passage à l'échelle avec un nombre limité d'échantillons de référence supplémentaires. Cette étude a débouché sur la création d'une chaîne de traitement opérationnelle permettant de produire la première carte des plantations de peuplier à l'échelle nationale à partir d'images satellitaires. Elle s'appuie sur un nouvel indice spectral proposé – le *Poplar Index (PI)* – qui exploite les bandes du *SWIR* et du *Red edge* des données Sentinel-2. Cet indice, et son évolution annuelle, ont permis de reconnaître les peupleraies avec une précision producteur de près de 95%. Le résultat de ce travail offre à la filière populicole une méthode robuste pour assurer une production annuelle d'une carte des peupleraies avec un niveau de fiabilité adapté.

Mots-clés : Peuplier ; Sentinel-2 ; Large échelle ; Adaptation de domaine ; Apprentissage automatique ; Classification ; Télédétection ; Cartographie

Abstract

Cultivated poplar is the most widely planted deciduous species in France. It plays an essential economic role, particularly in producing wood materials, plywood, light packaging and in the paper industry. However, despite the sector's importance, the national surface area of poplar plantations is highly uncertain. Depending on the data source considered (National forest database, land register, forest inventory), the estimates reveal differences of more than 50,000 ha. The updating rate of these data sources is unsuitable for monitoring cultivated poplars due to their short rotation cycle (15-20 years).

Satellite Remote Sensing has long been used for mapping forests, both natural and planted. With the continuous improvement of sensor characteristics, including spatial, spectral and temporal resolutions, it becomes possible to consider their application in an operational context for regular monitoring of the resource over large areas.

The aim of this thesis is twofold. The first one is to explore the potential of Sentinel-2 optical image time series to automatically distinguish poplar plantations from the other deciduous species while considering the diversity of poplar growing contexts. The second consists of developing a national classification strategy, taking into account the spatial non-stationarity of the spectral response of poplar plantations, the heterogeneity of the acquisitions, and the limited number of reference data.

The approach adopted consisted in investigating different domain adaptation techniques drawn from the field of machine learning. The unsupervised or semi-supervised techniques involved have made it possible to meet the large scale challenges while using a limited number of additional reference samples.

This study led to the development of an operational processing chain allowing the first map of poplar plantations on a national scale to be produced from satellite images. It is based on a newly proposed spectral index – the *Poplar Index (PI)* – exploiting the SWIR and Red edge bands of the Sentinel-2 images. Together with its annual evolution, this spectral index has enabled the identification of poplar plantations with a producer's accuracy of around 95%. The result of this work provides the poplar sector with a robust method to produce an annual map of poplar plantations with an appropriate level of reliability.

Mots-clés : Poplar ; Sentinel-2 ; Large scale ; Domain adaptation ; Machine learning ; Classification ; Remote sensing ; Mapping

Remerciements

“Il faut toujours remercier l’arbre à karité sous lequel on a ramassé de bons fruits pendant la bonne saison.”

— Ahmadou Kourouma

Je tiens à exprimer ma reconnaissance à toutes les personnes qui ont contribué de près ou de loin à l’aboutissement de ces travaux de thèse. Je remercie tout d’abord l’UMR DYNAFOR et le CNP qui ont été à l’origine de cette aventure CIFRE ainsi que tous les financeurs du secteur public et privé qui ont supporté cette thèse.

J’adresse en particulier mes premiers remerciements à David Sheeren, mon encadrant de thèse. Merci David pour tous tes précieux conseils et tes idées originales sans lesquels le travail ne serait pas ce qu’il est aujourd’hui. Merci pour ton encadrement, ton soutien et ta confiance grandissante au cours de ces années de thèse. Merci de t’être rendu toujours disponible dès que j’en avais besoin. Après nos discussions, je suis toujours repartie pleine d’énergie et de motivation. Tu n’as pas cessé de me booster et me tirer vers le haut en m’accordant toujours plus de confiance et j’ai appris à tes côtés ce qu’est la rigueur dans la recherche.

Je tiens à exprimer ma gratitude envers Eric Paillassa mon encadrant du côté de la filière peuplier. Je te remercie Eric pour ta bonne humeur permanente, ta confiance en mon travail et en son issue. Je te remercie également pour avoir été un excellent interlocuteur entre les télédéTECTEURS d’une part et les industriels et les forestiers d’autre part. Merci pour ton expertise thématique, tu m’as beaucoup appris sur les forêts et notamment les peupliers.

Je ne saurais poursuivre mes remerciements autrement qu’en les adressant à mes deux directeurs de thèse Claude Monteil et Véronique Chéret. Merci Claude pour ta présence et ton implication dans cette thèse. Tu as toujours été attentif lors de nos échanges et tu m’as toujours aidée à prendre du recul et à valoriser mes travaux. Merci d’avoir toujours répondu présent à chaque fois que je t’ai sollicité. J’ai beaucoup apprécié nos petits rendez-vous sur Zoom lors des deux confinements, c’était toujours très agréable et motivant ! Véronique, je te remercie pour ta présence, tes conseils scientifiques, mais aussi personnels ainsi que ta bienveillance. Tu m’as accompagnée avant même que je ne commence la thèse et tu as toujours été impliquée et soucieuse à mon égard. Merci pour ton attention particulière, tes encouragements et la confiance que tu m’as témoignée depuis que nous nous sommes rencontrées.

Je remercie chaleureusement l'ancien directeur de l'UMR DYNAFOR, Marc Deconchat, et les deux directrices actuelles, Aude Vialatte et Clélia Sirami, de m'avoir accueillie au sein du laboratoire. Merci pour votre bienveillance et votre soutien dans les moments les plus difficiles. Je tiens à remercier également Gabriel Chazallon, le président du CNP, de son accueil, sa confiance et sa patience face à des contraintes administratives pas toujours très simples, mais qu'il a toujours su gérer à temps.

Par ailleurs, je souhaite remercier les membres de mon comité de thèse, composé de Jordi Inglada, Pierre-Louis Frison et Jean-Baptiste Féret qui ont suivi ma thèse avec intérêt et m'ont fait part de leurs remarques et conseils avisés. De même, je tiens à remercier les membres du jury : Philippe Lejeune, Agnès Bégué, Thomas Corpetti et Xavier Briottet pour l'intérêt qu'ils ont accordé à mes travaux de thèse. Je vous remercie pour vos commentaires et suggestions ainsi que pour la riche discussion que nous avons eue pendant la soutenance. Elle m'a permis de prendre plus de recul et de porter un nouveau regard sur l'ensemble de mes travaux.

Je tiens également à remercier mes collègues du CESBIO pour toutes les interactions que nous avons eues. Je remercie particulièrement Benjamin Tardy et Arthur Vincent de m'avoir facilité la prise en main de *iota2*, de leur réactivité et de leur bonne humeur face aux bugs que nous avons rencontrés. Je remercie aussi Mathieu Fauvel et Vincent Thierion pour les différents échanges que nous avons eus et pour leurs disponibilités. De même, je remercie les collègues du CNES de m'avoir accordé l'accès à leurs ressources informatiques qui m'ont permis de lancer mes calculs à l'échelle nationale.

Je souhaite remercier les techniciens du CRPF que j'ai rencontrés au cours de cette aventure. Je pense particulièrement à Johann Hübelé et Nicolas Vanderheeren. Je vous remercie pour le partage des données terrain, les balades dans les peupleraies et les différentes connaissances thématiques que vous m'avez apportées. Je tiens également à remercier les popuiculteurs, les industriels et tous les autres membres du CNP qui m'ont toujours réservé un accueil chaleureux lors des différents événements avec la filière. Je remercie en particulier Yannick Bouraud, Bernard Mourlan, Sabine Boury et Emmanuel Naudin.

Je ne pourrais oublier d'inclure le collectif des non-permanents sur les trois sites de l'ENSAT, de INRAE et de Purpan. Merci à tous de votre présence, votre bonne humeur et les diverses discussions scientifiques et personnelles que nous avons partagées ! Ensemble, nous avons toujours réussi à célébrer nos succès, mais aussi à adoucir les périodes les plus difficiles. Un grand merci à Justine, James, Marc, Maïlys, Thierry, Florent, Mélodie, Rafaël, Nirina, Alexis, Clémence, Axelle, Gabriel, Théo, Estelle, Eve, Johanna, Fabio, Floriane et tous les autres qui ont croisé mon chemin. Je tiens à remercier les deux stagiaires Simon et Olivia pour l'intérêt qu'ils ont manifesté pour ma thèse et tout le travail qu'ils ont accompli.

Je remercie tous les membres de DYNAFOR, sans exception, je ne saurais citer tout le monde, mais c'est votre bienveillance et votre bonne ambiance qui rendent l'environnement de travail plus qu'agréable. Je garderai avec moi toute cette richesse et les beaux moments

que nous avons partagés ensemble. Je tiens à remercier tout particulièrement Richard pour son support informatique et sa bonne humeur, un grand merci à Sylvie pour son dynamisme et son beau sourire, merci aux gestionnaires d'unité pour leur efficacité et leur professionnalisme, avec une mention spéciale pour Valérie Calvo. Merci également aux autres collègues Ensatiens et notamment Annie, Bruno, Alzira, Betty, Loïc et Cécile pour toutes les pauses et les repas de fêtes que nous avons partagés mais aussi les cours de Yoga ! Un grand merci aux collègues Purpanais pour leur accueil et tous les échanges que nous avons pu avoir. Je pense particulièrement à Jean-Philippe, Asma, Marie et Anne.

Je tiens à remercier tous les membres de l'association ParenThèse Occitanie et Paca, avec qui j'ai partagé des journées entières de rédaction sur Zoom. Votre présence était une grande source d'inspiration et de motivation lors des deux confinements.

Je vais enfin conclure par exprimer toute ma gratitude et ma reconnaissance à ma famille qui m'a encouragée tout au long de mon parcours. Merci de m'avoir toujours soutenue dans mes différents choix et d'avoir toujours cru en moi. Merci de votre présence inconditionnelle malgré la distance qui nous sépare.

Je ne saurais terminer ces remerciements autrement que par mon mari, Amine. Tu m'as encouragée, tu as cru en moi et tu m'as toujours donné la confiance dont j'avais besoin pour aller de l'avant. Armé de patience, tu as été d'un soutien sans faille et tu as toujours su me reconforter dans les moments les plus difficiles. Je te remercie d'avoir enduré mes impatiences et je te remercie pour tous les sacrifices que tu as faits et les trajets hebdomadaires Toulouse - Montpellier que nous avons partagés ces dernières années.

Enfin, à toi mon petit bout d'chou, j'espère qu'un jour tu liras cette dédicace spéciale pour toi, merci d'avoir été plus ou moins sage et d'avoir accompagné maman pour la fin de sa rédaction et lors de sa soutenance !

Avant-propos

Cette thèse a été réalisée dans le cadre du dispositif Conventions Industrielles de Formation par la Recherche (CIFRE) numéro 2017/0228 de l'Association Nationale de la Recherche et de la Technologie (ANRT) associant le Conseil National du Peuplier (CNP) et l'unité mixte de recherche Dynamiques et Écologie des paysages Agriforestiers (DYNAFOR).

Le projet a bénéficié de financements publics et privés de la part de :

- Fonds propres du Centre National de la Propriété Forestière (CNP) et de l'UMR DYNAFOR.
- Ministère de l'Agriculture et de l'Alimentation.
- Bourse CIFRE de l'ANRT.
- Comité Professionnel de Développement Economique (CODIFAB)
- Conseil régional de Nouvelle-Aquitaine.
- France Bois Forêt.
- Alliance Forêts Bois.
- Conseil départemental de Lot-et-Garonne.
- Région Grand Est.
- Garnica Plywood.



Liste des contributions

Articles dans des revues internationales à comité de lecture

- **Hamrouni, Y.**, Paillassa, E., Chéret, V., Monteil, C., Sheeren, D. Sentinel-2 Poplar Index for Operational Mapping of Poplar Plantations over Large Areas. *Remote Sensing of Environment*. Article soumis pour publication le 19/06/2021.
- **Hamrouni, Y.**, Paillassa, E., Chéret, V., Monteil, C., Sheeren, D. From Local to Global : A Transfer Learning-Based Approach for Mapping Poplar Plantations at National Scale using Sentinel-2. *ISPRS Journal of Photogrammetry and Remote Sensing*, 171 (2021), pp. 76-100. <https://doi.org/10.1016/j.isprsjprs.2020.10.018>

Conférences internationales

Communications orales :

- **Hamrouni, Y.**, Paillassa, E., Chéret, V., Monteil, C., Sheeren, D. Operational Mapping of Poplar Plantations at the French National Scale Using Remote Sensing Satellite Images. *26th session of the International Commission on Poplars and Other Fast-Growing Trees Sustaining People and the Environment (IPC)*, Octobre 2021, FAO Rome, Italie.
- **Hamrouni, Y.**, Sheeren, D. Domain Adaptation by Optimal Transport for mapping poplar plantations over large areas. *IGARSS, Classification and Clustering of Satellite Image Time Series Session*, Juillet 2021, Bruxelles, Belgique.
- **Hamrouni, Y.**, Paillassa, E., Chéret, V., Monteil, C., Sheeren, D. Poplar plantations monitoring with Sentinels products : combination of optical and radar remote sensing data. *THEIA Workshop for Sentinel-2 L2A MAJA products*, Juin 2018, Toulouse, France

Posters :

- **Hamrouni, Y.**, Paillassa, E., Chéret, V., Monteil, C., Sheeren, D. From local to global : a transfer learning-based approach for mapping poplar plantations at large scale.

Mediterranean and Middle-East Geoscience and Remote Sensing Symposium, 9-11 Mars 2020, Tunisie.

- **Hamrouni, Y.**, Paillassa, E., Chéret, V., Monteil, C., Sheeren, D. Active Learning For Large-scale Classification of Poplar Plantations With Sentinels Time Series. *Living Planet Symposium, Forest Monitoring Session*, Mai 2019, Milan, Italie. https://github.com/Yousrah/LPS_YHamrouni/blob/master/LPS_2019_YHamrouni_HR.pdf

10.1109/IGARSS47720.2021.9554509

Articles de conférence :

- **Hamrouni, Y.**, Sheeren, D. Spatial and temporal domain adaptation by optimal transport for mapping poplar plantations over large areas. *IGARSS 2021 - IEEE International Geoscience and Remote Sensing Symposium*, 2021. <https://doi.org/10.1109/IGARSS47720.2021.9554509>.
- **Hamrouni, Y.**, Paillassa, E., Chéret, V., Monteil, C., Sheeren, D. Synergistic use of Sentinel-1 and Sentinel-2 time series for poplar plantations monitoring at large scale. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLIII-B3-2020, 1457–1461. <https://doi.org/10.5194/isprs-archives-XLIII-B3-2020-1457-2020>.
- **Hamrouni, Y.**, Paillassa, E., Chéret, V., Monteil, C., Sheeren, D. From Local to Global : A Transfer Learning-Based Approach for Mapping Poplar Plantations at Large Scale, *2020 Mediterranean and Middle-East Geoscience and Remote Sensing Symposium (M2GARSS)*, 2020, pp. 242-245. <https://doi.org/10.1109/M2GARSS47143.2020.9105218>.

Conférences nationales

- **Hamrouni, Y.**, Bernardoff, O., Karasiak, N., Sheeren, D. Adaptation de domaine par transport optimal pour la cartographie à large échelle des peupleraies. *SAGEO*, Mai 2021, La Rochelle, France.
- **Hamrouni, Y.**, Paillassa, E., Chéret, V., Monteil, C., Sheeren, D. Cartographie à large échelle des peupleraies à partir de séries temporelles d'images Sentinel-2. *SAGEO*, Novembre 2019, Clermont Ferrand, France. <https://hal.archives-ouvertes.fr/hal-02396759/document>.

Vulgarisations scientifiques

Communications orales :

- **Hamrouni, Y.**, Paillassa, Sheeren, D. Apport des séries temporelles d'images satellitaires pour caractériser la composition des peuplements forestiers. Français. *Les Agro-webinaires du mardi*, Mars 2021.
- **Hamrouni, Y.**, Paillassa, E., Chéret, V., Monteil, C., Sheeren, D. Identification et suivi des peupleraies par télédétection hypertemporelle. *Webinaire télédétection forestière*, Juin 2020.
- **Hamrouni, Y.** Ma thèse en 18 minutes à la Radio. Identification des peupleraies par suivi satellitaire. *Radio Campus FM Toulouse*, Décembre 2018.

Communications orales auprès de la filière peuplier :

- **Hamrouni, Y.**, Paillassa, E., Chéret, V., Monteil, C., Sheeren, D. Identification et suivi des peupleraies par télédétection hypertemporelle. Réunion de l'association Peuplier Nouvelle Aquitaine. Saintes, Juin 2019.
- **Hamrouni, Y.**, Paillassa, E., Chéret, V., Monteil, C., Sheeren, D. Identification et suivi des peupleraies par télédétection hypertemporelle. Assemblée générale de la Chambre du Peuplier. Arc-et-Senans, Avril 2019.
- **Hamrouni, Y.**, Paillassa, E., Chéret, V., Monteil, C., Sheeren, D. Identification et suivi des peupleraies par télédétection hypertemporelle. Groupe de travail Peuplier. Paris, Janvier 2019.
- **Hamrouni, Y.**, Paillassa, E., Chéret, V., Monteil, C., Sheeren, D. Identification et suivi des peupleraies par télédétection hypertemporelle. Assemblée générale du Conseil National du Peuplier. Paris, Décembre 2018.
- **Hamrouni, Y.**, Paillassa, E., Chéret, V., Monteil, C., Sheeren, D. Apport de la télédétection satellitaire pour le suivi des milieux forestiers : applications. InterCetef National. Bazas, Septembre. 2018.
- **Hamrouni, Y.**, Paillassa, E., Chéret, V., Monteil, C., Sheeren, D. Application de la télédétection satellitaire pour l'identification et le suivi des peupleraies. Réunion CRPF Occitanie. Montauban, Septembre 2018.

Articles :

- **Hamrouni, Y.** Identification et suivi des peupleraies par télédétection hypertemporelle. *Forêt Entreprise* 247. Août 2019, p. 31-36.



Activités d'encadrement

- O. Bernardoff, 2020. Stage de fin d'études de Master 2 (SIGMA) : Apprentissage par transfert pour la reconnaissance d'essences forestières à large échelle par télédétection. Co-encadrement avec D. Sheeren., N. Karasiak.
- S. Hano, 2019. Stage de fin d'études de Master 2 (Aménagement des territoires et télédétection) : Télédétection des changements dans les peupleraies. Co-encadrement avec D. Sheeren.

Activités d'enseignement

- Intervention dans l'Unité d'Enseignement Algorithmique et Modélisation Orientées Objet (AMOO). Cours Ingénieur 2A ENSAT. 12h ETD.
- Intervention dans l'Unité d'Enseignement Géomatique avancée : SIG, télédétection et modélisation. Initiation à la télédétection Radar. Master 2 SIGMA Université Toulouse Jean Jaurès/ENSAT. 4h ETD.
- Intervention dans le cadre de l'Unité d'Enseignement La recherche en géomatique. Master 2 SIGMA Université Toulouse Jean Jaurès/ENSAT. 4h ETD.
- Intervention dans le module SIG. Master 2 Sciences de l'Eau et de l'Environnement (SEE) ENSEEIHT. 8,75h ETD.
- Intervention dans l'Unité d'Enseignement Télédétection, SIG et Agriculture de Précision. Cours Ingénieur 1A ENSAT. 18h ETD.

Table des matières

| | |
|---|-----------|
| Résumé | ii |
| Abstract | iii |
| Remerciements | iv |
| Avant-propos | vii |
| Liste des contributions | viii |
| Table des matières | xv |
| Liste des figures | xx |
| Liste des tableaux | xxii |
| Liste des abréviations, des sigles et des acronymes | xxiii |
| | |
| I Introduction | 1 |
| | |
| Chapitre 1 : Contexte et problématique de la thèse | 2 |
| 1.1 Contexte de la recherche | 3 |
| 1.2 La télédétection pour le suivi des milieux forestiers | 13 |
| 1.3 La cartographie sur de grandes étendues par télédétection | 28 |
| 1.4 Positionnement de la thèse | 35 |
| | |
| II Matériels et méthodes | 39 |
| | |
| Chapitre 2 : Sites d'étude, données et pré-traitements | 40 |
| 2.1 Choix des sites populicoles | 41 |
| 2.2 Données satellitaires | 43 |

| | | |
|---|--|------------|
| 2.3 | Données de référence | 50 |
| 2.4 | Stratégie d'échantillonnage | 52 |
| Chapitre 3 : Classification supervisée à partir de séries temporelles d'images satellitaires | | 55 |
| 3.1 | Apprentissage automatique | 56 |
| 3.2 | Choix de l'algorithme de classification | 61 |
| 3.3 | Évaluation des performances | 64 |
| III Adaptation de domaine pour la classification à large échelle des peupleraies | | 68 |
| Chapitre 4 : Adaptation de domaine par transport optimal | | 69 |
| 4.1 | Principe du transport optimal | 70 |
| 4.2 | Application du transport optimal pour l'identification des peupleraies sur de grandes étendues | 80 |
| 4.3 | Conclusion | 107 |
| 4.4 | Bilan | 109 |
| Chapitre 5 : Adaptation de domaine par apprentissage actif | | 110 |
| 5.1 | Résumé | 111 |
| 5.2 | From local to global: A transfer learning-based approach for mapping poplar plantations at national scale using Sentinel-2 | 112 |
| 5.3 | Bilan | 144 |
| IV Vers une classification opérationnelle des peupleraies à l'échelle nationale | | 145 |
| Chapitre 6 : L'indice du peuplier pour la cartographie opérationnelle des peupleraies | | 146 |
| 6.1 | Résumé | 147 |
| 6.2 | Sentinel-2 Poplar Index for Operational Mapping of Poplar Plantations over Large Areas | 148 |
| 6.3 | Bilan | 176 |
| Chapitre 7 : Production à l'échelle nationale | | 177 |
| 7.1 | Projet TOSCA PARCELLE | 178 |
| 7.2 | Présentation de la chaîne de traitement iota^2 | 178 |

| | | |
|----------|--|------------|
| 7.3 | Présentation de l'infrastructure de calcul haute performance du CNES | 180 |
| 7.4 | Utilisation de iota^2 pour la production de la carte nationale des peupleraies sur le cluster HPC du CNES | 181 |
| V | Conclusion | 189 |
| | Chapitre 8 : Discussion | 190 |
| 8.1 | Synthèse des principales contributions | 191 |
| 8.2 | Qualité des échantillons de référence | 194 |
| 8.3 | Production cartographique | 197 |
| 8.4 | Stabilité de la production nationale | 203 |
| | Chapitre 9 : Conclusion générale | 205 |
| 9.1 | Conclusions | 206 |
| 9.2 | Perspectives | 207 |
| | Références bibliographiques | 209 |
| | Annexes | 229 |
| | Annexe A : Compléments chapitre 5 | 230 |
| A.1 | Acquisition dates of the Sentinel-2 images | 231 |
| A.2 | Example of the gap-filling process | 232 |
| A.3 | Classification results before the Active Learning adaptation | 233 |
| A.4 | Distribution of NDVI values within the deciduous classes of each tile | 235 |
| A.5 | Active learning vs. random sampling: class-specific assessment according to producer's and user's accuracy measures | 237 |
| A.6 | Active learning results with a combination of uncertainty and diversity measures | 239 |
| A.7 | Complete prediction maps | 240 |
| | Annexe B : Compléments chapitre 6 | 242 |
| B.1 | Resampled dates for 2017 and 2018 time series | 243 |
| B.2 | Correlation between the poplar index (PI) variants | 244 |
| B.3 | Additional spectral indices tested | 245 |
| B.4 | McNemar test p-values | 246 |
| B.5 | Spectral signatures | 248 |

Annexe C : Compléments chapitre 7 **250**
C.1 Acquisitions Sentinel-2 utilisées en 2018 251

Liste des figures

| | | |
|------|---|----|
| 1.1 | Spécimens d'herbier du genre <i>Populus</i> | 5 |
| 1.2 | Deux sites populicoles en France | 8 |
| 1.3 | Évolution de la récolte et des surfaces reboisées en peuplier | 9 |
| 1.4 | Années d'acquisition et de diffusion de la BD Forêt V2 par département métropolitain | 10 |
| 1.5 | Extrait du plan cadastral sur le département de la Haute-Garonne issu du cadastre de 2016 | 12 |
| 1.6 | Modes d'acquisition des systèmes de télédétection. | 14 |
| 1.7 | Opacité atmosphérique (absorption gazeuse) en fonction de la longueur d'onde. La courbe ne tient pas compte des atténuations potentielles. | 15 |
| 1.8 | Illustration de la notion de résolution spatiale sur des extraits d'images satellitaires en composition colorée infrarouge couleur d'une peupleraie dans les Coteaux de Gascogne. | 16 |
| 1.9 | Exemple de signatures spectrales de trois surfaces naturelles | 16 |
| 1.10 | Évolution d'une peupleraie entre août et octobre 2019 | 17 |
| 1.11 | Interaction entre le rayonnement électromagnétique et la feuille selon le domaine spectral. | 18 |
| 1.12 | Carte heuristique des principales composantes considérées dans la littérature pour le suivi des essences forestiers. | 22 |
| 1.13 | Schéma de la structure de la thèse. | 38 |
| 2.1 | Localisation des sites d'étude | 41 |
| 2.2 | Carte des surfaces de peupleraies par commune dérivées du cadastre 2003 | 42 |
| 2.3 | Illustrations de peupleraies typiques dans les trois sites de référence | 43 |
| 2.4 | Résolutions spatiale (taille du pixel au sol) et spectrale (largeur de bande) des images Sentinel-2. | 44 |
| 2.5 | Illustration des tuiles Sentinel-2 et des trajectoires au sol des orbites relatives passant par la France. | 44 |

| | | |
|------|--|-----|
| 2.6 | Illustration des deux niveaux de traitements 1C et 2A. | 47 |
| 2.7 | Nombre de fois où les pixels Sentinel-2 sont vus en France pour l'année 2018. | 48 |
| 2.8 | Exemple de sélection de peupleraies pour la photo-interprétation sur un fond d'orthophotographies IRC de 2017. | 51 |
| 2.9 | Illustration de la règle des $1,5 \times \text{EL}$ | 53 |
| 3.1 | L'Intelligence Artificielle et ses dérivés. | 56 |
| 3.2 | Exemple simple d'un <i>clustering</i> | 57 |
| 3.3 | Processus de la classification supervisée. | 59 |
| 3.4 | Illustration du dilemme biais-variance. | 60 |
| 3.5 | Illustration du principe de la validation croisée | 64 |
| 4.1 | Illustration du problème de transport optimal selon la formulation de Monge. | 71 |
| 4.2 | Exemple de deux distributions de points dans un espace de caractéristiques pour la source et la cible. | 75 |
| 4.3 | Transport optimal des deux mesures de probabilité | 75 |
| 4.4 | Matrices des couples optimaux. | 77 |
| 4.5 | Couplage entre les éléments de la source et de la cible | 78 |
| 4.6 | Résultats des échantillons transportés selon les couples optimaux formés dans la Figure 4.5. | 78 |
| 4.7 | Schéma des quatre variantes du transport optimal envisagées. | 84 |
| 4.8 | Matrices de confusion calculées pour le scénario 1 de transfert spatial | 88 |
| 4.9 | Évolution des performances en fonction du nombre d'étiquettes cibles y^c rajou- tées dans la configuration du transport optimal semi-supervisé. | 89 |
| 4.10 | Profil temporel de la classe peuplier dans la source et dans la cible avec les séries temporelles de 2017 avant transport optimal. | 91 |
| 4.11 | Profil temporel de la classe peuplier dans la source et dans la cible avec les séries temporelles de 2017 après transport optimal avec l'EMD. | 93 |
| 4.12 | Matrices des couplages optimaux dans le cas du transport spatial avec le scé- nario 3. | 94 |
| 4.13 | Matrices de confusions calculées pour le scénario 3 de transfert spatial | 95 |
| 4.14 | Résultat cartographique avant et après application du transport optimal pour un transfert spatial selon le scénario 3. | 96 |
| 4.15 | Profil temporel de la classe peuplier dans la source et dans la cible avec les séries temporelles de 2017 et 2018 avant transport optimal. | 99 |
| 4.16 | Matrices des couplages optimaux dans le cas du transport temporel avec le scénario 3. | 100 |

| | | |
|------|---|-----|
| 4.17 | Extrait 1 du résultat cartographique avant et après application du transport optimal pour un transfert temporel selon le scénario 3. | 102 |
| 4.18 | Extrait 2 du résultat cartographique avant et après application du transport optimal pour un transfert temporel selon le scénario 3. | 103 |
| 4.19 | Profil temporel des deux classes peuplier et non-peuplier avant transport optimal avec la bande 11 | 105 |
| 4.20 | Matrices des couplages optimaux dans le cas du transfert spatio-temporel avec le scénario 3. | 106 |
| 5.1 | Outline of the study areas. | 118 |
| 5.2 | Sentinel-2 acquisitions available in 2017 and percent cloud cover/date | 119 |
| 5.3 | Acquisition years of aerial photographs per district in the forest database used as reference. | 119 |
| 5.4 | Flow chart of the proposed approach for the classification of large areas by active learning. | 121 |
| 5.5 | Changes in the average OA scores on the south-western target tile 31TCJ according to the additional number of training samples from the initial model of the north-eastern source tile 31UEQ. | 129 |
| 5.6 | Average F-score values per class obtained over 30 independent repetitions as a function of the number of added samples. | 131 |
| 5.7 | Changes in the average poplar F-score values between the north-eastern 31UEQ and south-western 31TCJ tiles according to the number of added samples. | 132 |
| 5.8 | Changes in the average chestnut F-score values on the south-western target tile 31TCJ according to the number of added samples. | 133 |
| 5.9 | Changes in the average F-score values on the south-western target tile 31TCJ for the locust class according to the number of additional samples. | 135 |
| 5.10 | Four map extracts zoomed on poplar plantations comparing the predictions achieved before and after adaptation by AL. | 136 |
| 5.11 | Kernel density of NDVI for poplars in the south-western and the north-eastern tiles. | 138 |
| 5.12 | Changes in the average F-score values on the south-western target tile 31TCJ for the closed deciduous forest class according to the number of additional samples. | 141 |
| 6.1 | Sentinel-2 image tiles used in the study | 152 |
| 6.2 | Single and multi-feature SFFS configurations | 155 |

| | | |
|------|--|-----|
| 6.3 | Spectral reflectance of poplar plantations versus deciduous tree species on four different dates in 2017 | 157 |
| 6.4 | Temporal profile of Band 11 reflectance in poplar plantations versus deciduous tree species in 2017. | 158 |
| 6.5 | Average poplar PA as a function of the number of features selected with SFFS in the multi-band configuration. | 161 |
| 6.6 | Average poplar PA as a function of the number of features selected with SFFS in the multi-index configuration. | 162 |
| 6.7 | Average poplar PA as a function of the number of features selected with SFFS in the single-band configuration. | 164 |
| 6.8 | Average poplar PA as a function of the number of features selected with SFFS in the single-index configuration. | 166 |
| 6.9 | Map extracts located in the Lot-et-Garonne department in tile 30TYQ. | 168 |
| 6.10 | Map extracts located in the Indre department in tile 31TCM. | 168 |
| 6.11 | Map extracts located in the Aquitaine department in tile 30TYQ. | 169 |
| 6.12 | Bivariate choropleth map of France showing the density of poplar plantations in each municipality with the median prediction confidence values for the year 2018. | 170 |
| 6.13 | Distribution of classification prediction confidence values as a function of the age of the detected plantations and of the cultivar planted | 173 |
| 7.1 | Illustration des sept principaux groupes formant le flux de travail dans iota ² | 180 |
| 7.2 | Emprises des 90 tuiles Sentinel-2 couvrant l'ensemble de la France selon les trois zones UTM. | 182 |
| 7.3 | Emprise de tuile 31UCQ selon trois orbites relatives. | 183 |
| 7.4 | Mosaïque de la classification des 90 tuiles Sentinel-2 à la sortie de la chaîne de traitement iota ² | 185 |
| 7.5 | Extraits cartographiques des classifications de peupleraies et des confiances associées sur trois départements métropolitains. | 187 |
| 8.1 | Polygones de peupleraies dérivés de la BD Forêt | 195 |
| 8.2 | Exemples de peupliers naturels détectés en bordure de cours d'eau | 199 |
| 8.3 | Illustration du masque forêt généré à partir de la moyenne temporelle hivernale des images Sentinel-1 | 201 |
| 8.4 | Carte choroplèthe bivariée de la France présentant la densité des peupleraies dans chaque commune avec les valeurs médianes de la confiance des prédictions pour l'année 2018. | 202 |

| | | |
|------|--|-----|
| 8.5 | Superposition des limites des parcelles cadastrales sur les prédictions en peupleraies. | 203 |
| A.1 | Example of the gap-filling process. | 232 |
| A.2 | Normalised confusion matrices obtained before active learning adaptation from the north-eastern source tile. | 233 |
| A.3 | Normalised confusion matrices obtained before active learning adaptation from the central source tile. | 233 |
| A.4 | Normalised confusion matrices obtained before active learning adaptation from the south-western source tile. | 234 |
| A.5 | Spectral variability of Sentinel-2 NDVI over the deciduous classes in the north-eastern, central and south-western tiles. | 235 |
| A.6 | Spectral variability of Sentinel-2 NDVI between locusts and closed deciduous forests on the north-eastern and south-western tiles. | 236 |
| A.7 | Kernel density of NDVI for locusts and closed deciduous forest in the north-eastern and south-western tiles. | 236 |
| A.8 | Average producer's accuracy per class as a function of the number of added samples. | 237 |
| A.9 | Average user's accuracy per class as a function of the number of added samples. | 238 |
| A.10 | Prediction maps according to the three-tile global model built by AL. | 241 |
| B.1 | Spearman's ρ correlation coefficients between poplar index (PI) variants based on all resampled dates for 2017 and 2018 S2 time series. | 244 |
| B.2 | Temporal profile of band 5 reflectance in poplar plantations versus other deciduous tree species. | 248 |
| B.3 | Temporal profile of PI_2 reflectance in poplar plantations versus other deciduous tree species. | 249 |

Liste des tableaux

- 2.1 Classes de feuillus extraites sur les trois sites d'étude. 51
- 2.2 Nombre total des échantillons de référence par classe extraits pour chaque tuile d'étude. 54
- 4.1 Temps de parcours entre les peupleraies (*i*) et les industries de transformation (*j*). 71
- 4.2 Nombre d'échantillons par classe selon le scénario considéré dans le transfert spatial. 86
- 4.3 Résultats du transfert spatial avec les quatre configurations de transport optimal. 87
- 4.4 Nombre d'échantillons par classe considéré pour le transfert temporel. 97
- 4.5 Résultats du transfert temporel avec les quatre configurations de transport optimal. 98
- 4.6 Nombre d'échantillons par classe considéré pour le transfert spatio-temporel. . 101
- 4.7 Résultats du transfert spatio-temporel avec les quatre configurations de transport optimal. 104
- 5.1 Total number of reference samples per class extracted from each S2 tile. 122
- 5.2 Results of site-specific classifications for each S2 tile averaged over 30 independent repetitions. 127
- 5.3 Average overall accuracy values according to the transfer direction and the number of target samples. 128
- 5.4 Average overall accuracy values according to the transfer direction from a two-tiles based model to a third one. 130
- 5.5 Average locust F-score on the south-western tile according to the total number of additional samples with and without the presence of mixed classes in the training set. 134
- 6.1 Total number of reference samples per class extracted from each S2 tile 154

| | | |
|-----|--|-----|
| 6.2 | List of the spectral indices compared with the corresponding formulas and references. | 159 |
| 6.3 | Poplar PA scores obtained using the 10 best feature subsets based on SFFS in the single and multi-feature configurations. | 167 |
| 7.1 | Caractéristiques et ressources accessibles sur le cluster HPC | 181 |
| A.1 | Acquisition dates in each Sentinel-2 tile over 2017. | 231 |
| A.2 | Overall accuracy values according to the transfer direction and the number of target samples using uncertainty and diversity metrics. | 239 |
| B.1 | Common resampled dates over 2017 and 2018 time series. | 243 |
| B.2 | List of additional spectral indices tested and their corresponding formulas. | 245 |
| B.3 | Classification comparison with the p-values derived from the McNemar significance test of performance on the test set. | 246 |
| B.4 | Poplar PA scores achieved with the ten best feature subsets based on SFFS in the single-feature configuration using the three variants of the poplar index | 247 |
| C.1 | Nombre d’acquisitions Sentinel-2 disponibles par tuile en 2018. | 251 |

Liste des abréviations, des sigles et des acronymes

| | |
|---------------------|--|
| ANRT | Association Nationale de la Recherche et de la Technologie. vii |
| BP | Bois de Peuplier. 11 |
| CES OSO | Centre d'Expertise Scientifique de l'Occupation des SOs. 31 , 49 , 173 , 178 , 182 , 185 , 197 , 201 |
| CESBIO | Centre d'Études Spatiales de la BIOSphère. 44 , 175 , 178 |
| Chloroplaste | organites présent dans les cellules végétales permettant de capter la lumière nécessaire à la photosynthèse. Il renferme principalement les pigments foliaires chlorophylliens.. 18 |
| CIFRE | Conventions Industrielles de Formation par la Recherche. vii |
| CIP | Commission Internationale du Peuplier. 4 , 6 , 8 |
| CNES | Centre National d'Études Spatiales. 44 , 177 , 178 , 181 |
| CNP | Conseil National du Peuplier. vii , 8 , 9 , 185 , 201 |
| Cultivar | abréviation issue du terme <i>Cultivated Variety</i> . Elle désigne une variété d'une espèce obtenue par sélection. 7 |
| DLR | <i>German Aerospace Center</i> . 44 |
| DLT | <i>Dominant Leaf Type</i> . 183 |
| DTW | <i>Dynamic Time Warping</i> . 107 |
| DYNAFOR | Dynamiques et Écologie des paysages Agriforestiers. vii , 175 |
| EEA | <i>European Environment Agency</i> . 183 , 198 , 199 |
| EI | Écart Interquartile. 52 , 181 |
| EMD | <i>Earth Mover's Distance</i> . 75–78 , 82 , 85 , 86 , 89 , 92 , 102 , 103 |
| EnMAP | <i>Environmental Mapping and Analysis Program</i> . 20 |
| EOS | <i>Earth Observing System</i> . 19 |

| | |
|-------------------------|--|
| ESA | <i>European Space Agency.</i> 20, 42, 43 |
| ETM + | <i>Enhanced Thematic Mapper Plus.</i> 23 |
| FAO | <i>Food and Agriculture Organization of the United Nations.</i> 4, 183 |
| FRE | <i>Flat REflectance.</i> 45, 179 |
| GPFS | <i>General Parallel File System.</i> 178 |
| GRECO | <i>Grandes Régions ÉCOlogiques.</i> 40, 41 |
| HPC | <i>High Performance Computing.</i> 177 |
| HRL | <i>High Resolution Layer.</i> 95, 100, 101, 183, 198, 199 |
| HSI | <i>Hyperspectral Imager.</i> 20 |
| IA | <i>Intelligence Artificielle.</i> 55 |
| IFN | <i>Inventaire Forestier National.</i> 9, 40, 41 |
| IGN | <i>Institut national de l'information géographique et forestière.</i> 9, 11 |
| iota² | <i>Infrastructure pour l'Occupation des sols par Traitement Automatique Incorporant les Orfeo Toolbox Applications.</i> 173, 175–178, 180–183, 190, 200, 205 |
| IPC | <i>International Poplar Commission.</i> 4 |
| IRC | <i>infrarouge couleur.</i> 10, 50 |
| LIDAR | <i>Light Detection And RAnging.</i> 14, 22, 23 |
| MACCS | <i>Multi-sensor Atmospheric Correction and Cloud Screening.</i> 44 |
| MGRS | <i>Military Grid Reference System.</i> 43, 179 |
| MNT | <i>Modèle Numérique de Terrain.</i> 43, 45 |
| MODIS | <i>Moderate Resolution Imaging Spectroradiometer.</i> 19, 23, 25, 27, 28, 33 |
| MSI | <i>MultiSpectral Instrument.</i> 42 |
| MSS | <i>Multispectral Scanner System.</i> 19 |
| NASA | <i>National Aeronautics and Space Administration.</i> 19 |
| OA | <i>Overall Accuracy.</i> 64–66, 84–88, 97, 102, 103, 181 |
| OLI | <i>Operational Land Imager.</i> 20 |
| PA | <i>Producer's Accuracy.</i> 65, 66, 84–89, 92, 97, 102, 103, 188, 189 |
| PARCELLE | <i>Production cARtographique grande éChELLE par tÉLédétEction.</i> 175, 177, 178 |
| Peupleraie | <i>plantation de peupliers.</i> 7 |



| | |
|-------------------------|---|
| Phytoremédiation | capacité des végétaux à réduire les composés polluants des sols, des eaux et de l'air. 6 |
| Populiculteur | gestionnaire forestier cultivant du peuplier. 8 |
| Populiculture | culture du peuplier. 7 |
| <i>Populus</i> | genre de la famille des <i>Salicaceae</i> regroupant toutes les espèces de peupliers. 3 |
| POT | <i>Python Optimal Transport</i> . 84 |
| PRISMA | <i>PRecursore IperSpettrale della Missione Applicativa</i> . 20 |
| RADAR | RAdio Detection And RAnging. 14 , 15 , 22 , 23 , 25 , 26 , 56 , 197 |
| REM | rayonnement électromagnétique. 13–15 , 17 |
| RF | <i>Random Forest</i> . 57 , 60–63 , 84 , 181 , 193 |
| Ripicole | caractéristique de la ripisylve, formations végétales qui occupent les rives des cours d'eau. 3 , 6 |
| RSO | Radar à Synthèse d'Ouverture. 15 |
| SER | Sylvoécorégions. 40 , 41 |
| SITS | <i>Satellite Image Time Series</i> . 116 , 119 , 121 , 125 |
| SPOT | Satellites Pour l'Observation de la Terre. 19 |
| SRE | <i>Surface REflectance</i> . 45 |
| SVM | <i>Support Vector Machine</i> . 57 , 60 |
| SWIR | <i>Short Wave Infrared</i> . 145 , 173 , 190 |
| TFV | Type de Formation Végétale. 50 |
| TIRS | <i>Thermal Infrared Sensor</i> . 20 |
| TM | <i>Thematic Mapper</i> . 23 |
| TOA | <i>Top Of Atmosphere</i> . 43 , 45 |
| TOC | <i>Top Of Canopy</i> . 44 , 45 |
| UA | <i>User's Accuracy</i> . 65 , 66 |
| USGS | <i>United States Geological Survey</i> . 20 |

Première partie

Introduction

Chapitre 1

Contexte et problématique de la thèse

Sommaire

| | | |
|-------|---|-----------|
| 1.1 | Contexte de la recherche | 3 |
| 1.1.1 | Le peuplier : arbre du XXI ^e siècle | 3 |
| 1.1.2 | Peuplier et populiculture en France | 6 |
| 1.1.3 | La surface des peupleraies en France : une grande incertitude | 9 |
| 1.2 | La télédétection pour le suivi des milieux forestiers | 13 |
| 1.2.1 | Généralités sur la télédétection spatiale | 13 |
| 1.2.2 | Contexte actuel : évolution des caractéristiques des capteurs de télédétection satellitaire | 19 |
| 1.2.3 | Télédétection forestière : un état de l'art | 21 |
| 1.3 | La cartographie sur de grandes étendues par télédétection | 28 |
| 1.3.1 | Les contraintes pour le passage à l'échelle | 28 |
| 1.3.2 | Approches adoptées pour la cartographie opérationnelle sur de grandes étendues | 30 |
| 1.4 | Positionnement de la thèse | 35 |
| 1.4.1 | Problématique et questions de recherche | 35 |
| 1.4.2 | Objectifs et démarche suivie | 36 |
| 1.4.3 | Organisation du manuscrit | 37 |

1.1 Contexte de la recherche

1.1.1 Le peuplier : arbre du XXI^e siècle

Les peupliers sont des arbres feuillus du genre *Populus* et forment avec les saules (genre *Salix*) la famille des Salicacées (*Salicaceae*) considérée par [Gordon \(2011\)](#) comme l'une des familles les plus importantes pour les activités humaines depuis l'aube des temps. Les deux Salicacées se distinguent entre eux par la forme de leurs feuilles qui sont généralement aussi longues que larges pour les peupliers et longues et étroites pour les saules ([Marie-Victorin et al., 2002](#)).

Les peupliers sont naturellement originaires des régions septentrionales et tempérées de l'hémisphère Nord et sont répartis principalement en Amérique du Nord, en Europe, en Asie occidentale et en Afrique du Nord. Le genre *Populus* est taxonomiquement divisé en six sections (Figure 1.1) comportant plus de 30 espèces ([Heilman, 1999](#)) :

Abaso : elle est représentée par l'unique espèce *Populus mexicana*. Elle est originaire du Mexique et des régions les plus chaudes du sud de l'Amérique du Nord et elle est considérée comme l'une des espèces les plus anciennes du genre *Populus* ([Eckenwalder, 1996](#); [Liu et al., 2017b](#)).

Aigeiros : c'est la section la plus répandue dans le monde ([Isebrands and Richardson, 2014](#)). Elle est représentée par deux principales espèces, les peupliers noirs *Populus nigra* et les peupliers deltoïdes *Populus deltoides*. Les peupliers noirs sont originaires de l'Eurasie et de l'Afrique du Nord alors que les deltoïdes sont naturellement distribués dans le sud du Canada et dans presque toute la moitié sud de l'Amérique du Nord allant des montagnes Rocheuses aux Grandes Plaines à l'Est et jusqu'au golfe du Mexique au Sud.

Leucoïdes : elle est composée de trois espèces localisées dans les régions chaudes et tempérées de l'Amérique du Nord et de l'Asie orientale. Les espèces de cette section sont réputées pour leur faible enracinement et la difficulté de leur culture à partir de boutures limitant ainsi l'intérêt porté à cette section ([Isebrands and Richardson, 2014](#)).

Populus : autrefois dénommée la section Leuce, elle regroupe les deux sous-sections des peupliers blancs (*Albidae*) et des Trembles (*Trepidae*). Tous les peupliers blancs appartiennent à l'espèce unique de *Populus alba*, une espèce principalement ripicole qui se trouve en Europe, en Asie et en Afrique du Nord. Les Trembles sont représentés par le *Populus tremula* eurasiatique et le *Populus tremuloides* américain et sont plutôt des espèces forestières ([Du Cros, 1984](#)).

Tacamahaca : c'est la section des peupliers baumiers originaires du nord de l'Amérique et de l'est de l'Asie. Elle comporte une dizaine d'espèces dont la plus répandue est *Populus*

balsamifera.

Turanga : elle est représentée par son espèce la plus répandue de peupliers d'Euphrate *Populus euphratica*. Elle est considérée comme l'espèce de peupliers la plus tenace du fait de son développement caractéristique dans les régions chaudes et sèches. Elle est naturellement distribuée de manière discontinue du Maroc et de l'Égypte en Afrique à l'Asie centrale et occidentale en passant par le Moyen-Orient (Isebrands and Richardson, 2014).

Les peupliers ont particulièrement suscité l'attention des anciennes civilisations grâce à leur croissance rapide et la facilité de leur propagation (Isebrands and Richardson, 2014) principalement dans les régions dépourvues de forêts naturelles de l'hémisphère Nord (Heilman, 1999). Ils ont été la principale source de bois utilisé pour la cuisson, le chauffage et la construction et ont même constitué des supports pour les peintres de la Renaissance et de la période impressionniste (Stettler, 2009).

Lors des explorations européennes des XVII^e et XVIII^e siècles, des boutures de peupliers de différentes espèces ont été ramenées pour être plantées dans des arboretums et des jardins. Parmi ces spécimens figurait le peuplier deltoïde (*Populus deltoides*) qui fut introduit en France depuis l'Amérique du Nord au début des années 1700 (Isebrands and Richardson, 2014). Compte tenu de leur prédisposition à l'hybridation, les fleurs des peupliers femelles nord-américaines apportées en Europe étaient réceptives au pollen du peuplier noir indigène (*Populus nigra*), générant des « hybrides spontanés » décrits pour la première fois en France en 1755 (Dickmann, 2006) et actuellement connus comme les peupliers euro-américains. La popularité de ces hybrides a motivé le développement de croisements contrôlés entre différents taxons de *Populus* à travers les travaux pionniers de Henry (1914) et Stout and Schreiner (1933). Plusieurs projets de sélection de peupliers ont ensuite vu le jour au Canada, aux États-Unis et dans de nombreux pays d'Europe et de Scandinavie visant à sélectionner des hybrides en fonction de leur productivité et de leur résistance au froid et aux maladies (Dickmann, 2006).

Après la seconde guerre mondiale, la demande croissante des produits du bois et l'augmentation de la population mondiale ont accéléré la diffusion et la plantation des hybrides de peuplier à croissance rapide à travers le monde. En Europe, sa plantation était considérée primordiale pour reconstituer les paysages dégradés et relancer l'économie dans les zones rurales (FAO, 2016b). Ceci a conduit à la création de la Commission Internationale du Peuplier (CIP) ou *International Poplar Commission* (IPC) en 1947 au sein de l'Organisation des Nations Unies pour l'alimentation et l'agriculture ou *Food and Agriculture Organization of the United Nations* (FAO) et qui regroupe actuellement 38 pays membres des cinq continents. Ses activités visent à promouvoir la plantation et la conservation des peupliers et à soutenir et co-



Populus mexicana (section Abaso)



Populus nigra (section Aigeiros)



Populus heterophylla (section Leucoides)



Populus alba (section Populus)



Populus balsamifera (section Tacamahaca)



Populus euphratica (section Turanga)

FIGURE 1.1 – Spécimens d'herbier du genre *Populus*. Source : [SEINet Portal Network](#) (2021).

ordonner les recherches relatives à sa culture, sa gestion et son utilisation au niveau mondial. Depuis 1980, la CIP a élargi son champs d'action pour inclure les saules et plus récemment, d'autres arbres à croissance rapide.

Aujourd'hui la majeure partie des peupliers naturels est partagée entre le Canada, la Russie et les États-Unis. Selon les dernières déclarations nationales à la 25^e édition de l'IPC réunie en 2016 (FAO, 2016a), les trois pays couvrent 96% de la surface mondiale des peupliers naturels estimée à plus de 50 millions d'hectares. Avec les progrès en amélioration génétique, des hybrides de plus en plus productifs et résistants sont plantés dans le monde entier non seulement dans des régions qui n'ont jamais connu le peuplier auparavant mais aussi dans les pays possédant des superficies importantes comme par exemple au sud de la Russie.

Alors que les réserves de pétrole continuent à s'épuiser et le prix du baril ne cesse d'augmenter, le bois de peuplier apparaît comme une matière première domestique et industrielle de plus en plus importante ainsi qu'une source alternative d'énergie (Dickmann, 2006). Ceci est parfois sujet à controverse dans différents pays du point de vue de son impact environnemental. Pour certains, les plantations de peupliers sont jugées préjudiciables à la biodiversité du fait de leur conduite en monoculture équienne qui offre moins d'opportunités et d'habitats pour les insectes et les oiseaux nicheurs, notamment à cause de l'absence de strates arbustives et buissonnantes (Marion Zapater et al., 2012). Pour d'autres, bien que cela soit fonction du mode de gestion sylvicole, les plantations de peupliers sont considérées plus riches en biodiversité que les parcelles agricoles conventionnelles, outre leurs atouts en termes d'amélioration des propriétés des sols et des eaux (Jug et al., 1999; Kahle et al., 2005).

Si la principale motivation pour la plantation de peupliers est économique et stimulée par l'intérêt de l'industrie forestière à sécuriser son approvisionnement en bois et sa qualité, une nouvelle tendance se dessine en faveur de leur utilisation à des fins environnementales, notamment pour la phytoremédiation et la réhabilitation des paysages. Ces applications environnementales et le besoin de ressources naturelles et renouvelables constituent désormais un enjeu majeur conférant souvent au peuplier le statut d'« arbre du XXI^e siècle » (Stobrawa, 2014).

1.1.2 Peuplier et popiculture en France

Les peupliers constituent depuis toujours un élément distinctif du paysage naturel de la France, et depuis plus d'un siècle de celui façonné par l'Homme. En France, il existe trois espèces endémiques de peuplier qui se sont naturellement développées : le peuplier noir, le peuplier blanc et le peuplier tremble. La principale espèce indigène est le peuplier noir (*Populus nigra*) qui est essentiellement ripicole et se développe dans les milieux alluviaux. Le peuplier blanc (*Populus alba*) est présent dans les vallées méditerranéennes et domine les ripisylves du

bassin méditerranéen. Le peuplier tremble (*Populus tremula*) est la seule espèce de peuplier forestier présente dans la plupart des massifs forestiers (Carah Asbl, 2019).

En France, la majeure partie des peupliers sont issus de plantations destinées à la production de bois d'œuvre. D'après les derniers chiffres annoncés à la CIP de 2016 (FAO, 2016a), leur surface est estimée à près de 212 000 ha. On trouve également des peupliers naturels qui occupent une surface plus réduite. Ils bordent principalement les rives des cours d'eau et forment, avec d'autres espèces, la ripisylve. Ils ont essentiellement une fonction de protection de l'environnement, notamment la stabilisation des berges, et une infime partie est dédiée à la production de bois ronds industriels ou de biomasse (FAO, 2016a).

Tous les peupliers plantés sont des cultivars issus d'hybridation. Grâce à leurs caractéristiques particulières combinant légèreté et résistance, ils sont entièrement destinés à la production de bois ronds industriels. Les principaux cultivars plantés en France sont des hybrides euraméricains et inter-américains mais aussi des représentants des deux espèces naturelles de *Populus trichocarpa* (section Tacamahaca) et *Populus deltoides* (section Aigeiros).

Les peupliers cultivés, dénommés peupleraies, ont une croissance rapide avec des cycles de 15 à 20 ans en moyenne. Ils se développent particulièrement bien sur des sols riches en éléments nutritifs et avec un apport suffisant en eau.

L'aspect de la peupleraie est fonction du mode de gestion sylvicole, qui est lui-même fonction des conditions stationnelles. Par exemple, dans les stations populicoles sèches, le travail du sol est nécessaire afin de limiter la compétition pour l'eau entre les arbres et le sous-étage alors que dans les régions plus humides et fraîches, l'installation d'une strate herbacée voire ligneuse est plus courante dans les peupleraies (Figure 1.2).

La populiculture alimente une filière locale qui garantit plusieurs emplois dans les territoires ruraux. Elle compte aujourd'hui environ 190 000 propriétaires cultivateurs de peupliers, ou populiculteurs, plus de 70 pépiniéristes qui élèvent et vendent environ un million de jeunes plants tous les ans, des industriels regroupant transformateurs de bois et fabricants d'emballages légers ainsi qu'une multitude d'acteurs du monde forestier incluant des experts, des exploitants, des transporteurs de grumes ou de bois, des associations et des coopératives forestières, etc.

La filière peuplier française est représentée par le CNP dont les principales missions sont d'animer la filière au niveau national et international et de promouvoir la plantation, la gestion et l'usage du peuplier. Le CNP arrime aujourd'hui tous les maillons de la filière populicole. La France fait partie des premiers pays membres de la CIP. Elle a accueilli en 1947 sa première session organisée par la Commission Nationale du Peuplier, devenue Conseil National du Peuplier depuis 2005.

En France, le volume de bois de peuplier récolté représentait 1,4 M m³ en 2017, soit 26,5%



(a) Peupleraies dans le département du Tarn-et-Garonne (82)

(b) Peupleraies dans le département de la Marne (51)

FIGURE 1.2 – Deux sites populicoles en France. En (a) une peupleraie dans le sud de le France avec un sous-étage très peu développé et en (b) une peupleraie dans la région Grand Est avec une strate herbacée et arbustive en sous-étage.

de la récolte de bois feuillus. En moyenne, 20% du volume récolté est exporté et le reste est consommé par les industries de transformation du bois et d’emballages légers. Les sous produits sont essentiellement valorisés dans l’industrie papetière, le paillage ainsi que dans la production d’énergie ([Conseil National du Peuplier, 2019](#)).

D’une manière générale, les parcelles récoltées sont renouvelées dans l’année qui suit leur exploitation. Cependant, depuis une vingtaine d’année les rythmes de récolte et de replantation ont subi des fluctuations importantes. À la fin des années 1990, malgré un rythme de reboisement élevé de plus de 9 000 ha par an, le volume récolté était encore plus important et se traduisait par un déficit annuel de replantation avoisinant les 1 000 ha. Si le taux de replantation a atteint un meilleur équilibre avec celui de la récolte entre 2002 et 2005, il s’est à nouveau dégradé entre 2006 et 2011 entraînant un déficit moyen de reboisement de 35% sur la période. Selon [Paillassa \(2014\)](#), cette baisse du taux de reboisement était principalement liée à la chute des prix du bois en raison de la crise économique, ce qui a rendu les populiculteurs moins enclins à renouveler les peupleraies exploitées.

Afin de relancer le rythme de replantation et de sécuriser les approvisionnements futurs en peuplier, de nombreux professionnels de la filière se mobilisent à travers des incitations financières au reboisement, notamment la charte « Merci le Peuplier », née en 2011 dans la région Ouest et étendue depuis 2014 à tout le territoire national. Bien que la hausse du taux

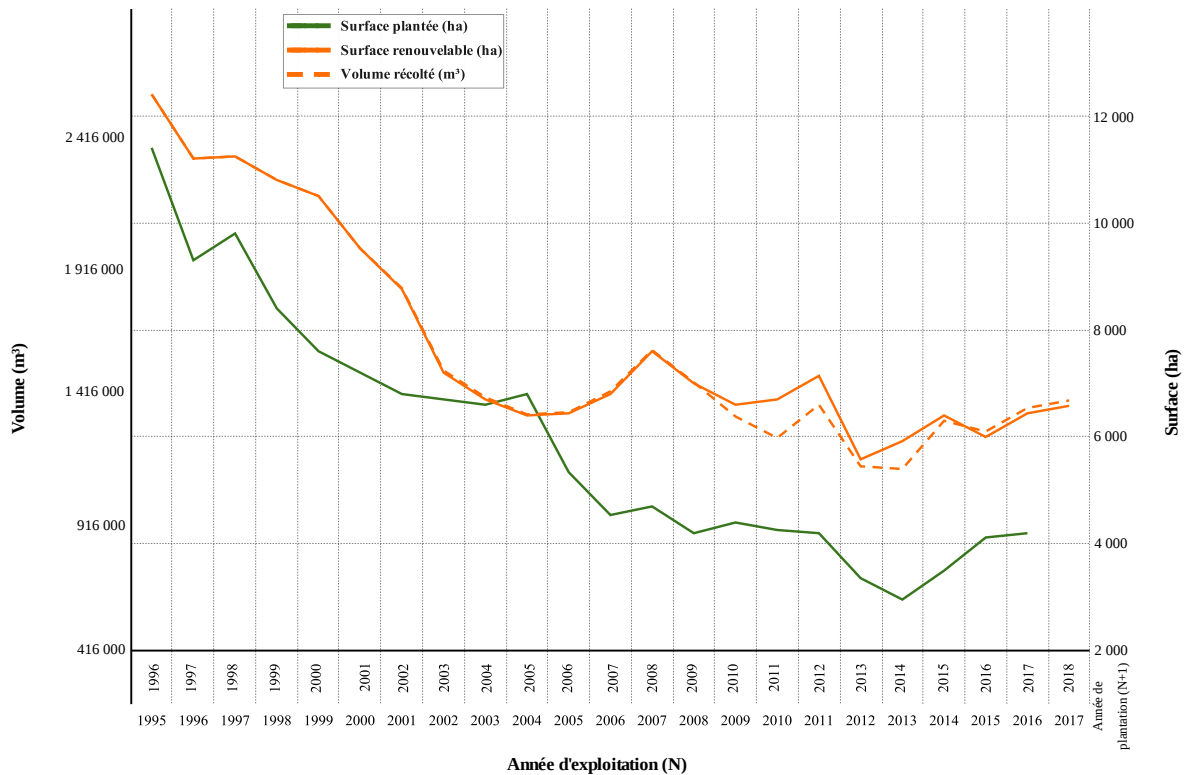


FIGURE 1.3 – Évolution de la récolte et des surfaces reboisées en peuplier. Source : adaptée depuis Conseil National du Peuplier (2019).

de reboisement entre 2014 et 2018 constitue un espoir pour la filière, les surfaces plantées demeurent bien en deçà des besoins et une pénurie de la ressource apparaît inévitable. Ce déficit est estimé par le CNP à 500 000 m³/an en 2025, soit environ 35% du volume exploité.

Face à une disponibilité inférieure au besoin des industries, et avant de pouvoir envisager un recours à l’importation ou à la substitution du bois de peuplier par celui d’autres essences, il convient d’abord d’évaluer la ressource disponible. Néanmoins, ce stock n’est pas connu avec précision et la filière se retrouve confrontée à une question récurrente sur la surface actuelle des peupleraies et son évolution dans le temps et dans l’espace, ce qui constitue une préoccupation majeure pour le CNP.

1.1.3 La surface des peupleraies en France : une grande incertitude

En France, la surface des peupleraies est principalement déterminée à partir de trois sources d’information : la BD Forêt, l’inventaire forestier et le cadastre.

La BD Forêt® :

La BD Forêt® est une base de données de référence qui décrit les formations végétales forestières et naturelles sur le territoire français. Elle a été d’abord développée par l’Inventaire Forestier National (IFN), qui en a initié une deuxième version « BD Forêt® version 2.0 » (V2)

en 2007. C'est désormais l'Institut national de l'information géographique et forestière (IGN) qui en assure la production et la diffusion depuis la fusion en 2012 de l'IFN avec l'ancien IGN (Institut Géographique National).

La BD Forêt V2 comporte une nomenclature nationale distinguant 32 types ou niveaux de formations végétales pour toutes les parcelles ayant une superficie d'au moins 0.5 hectares. L'approche adoptée pour sa création est basée sur la couverture du sol et la traduit en termes de densité du couvert, de composition et d'essences dominantes. Elle distingue ainsi des peuplements purs dont le taux de couverture est supérieur ou égal à 75% et des peuplements mixtes lorsque la couverture est inférieure à ce seuil (IGN, 2014). La BD Forêt V2 est élaborée à partir d'orthophotographies acquises entre 2004 et 2015 et elle est fournie progressivement deux à six ans après en fonction des différents millésimes pour chacun des 96 départements métropolitains. Avec tous les moyens conséquents déployés, la production d'une couverture nationale avec la BD Forêt a pris 11 ans et elle a été achevée en 2018 (Figure 1.4).

La BD Forêt est générée à partir d'un ensemble de traitements appliqués sur les orthophotographies infrarouge couleur (IRC) de la BD ORTHO®. La première étape consiste en une segmentation automatique permettant d'identifier les contours des différentes formations végétales. Les polygones issus de la segmentation sont ensuite classés automatiquement en cinq grandes classes en fonction du type de la couverture. Les experts forestiers interviennent ensuite pour identifier, par photo-interprétation, les essences forestières qui composent chaque polygone. Le processus est complété par une vérification sur le terrain afin de rectifier les éventuelles erreurs liées à la photo-interprétation.

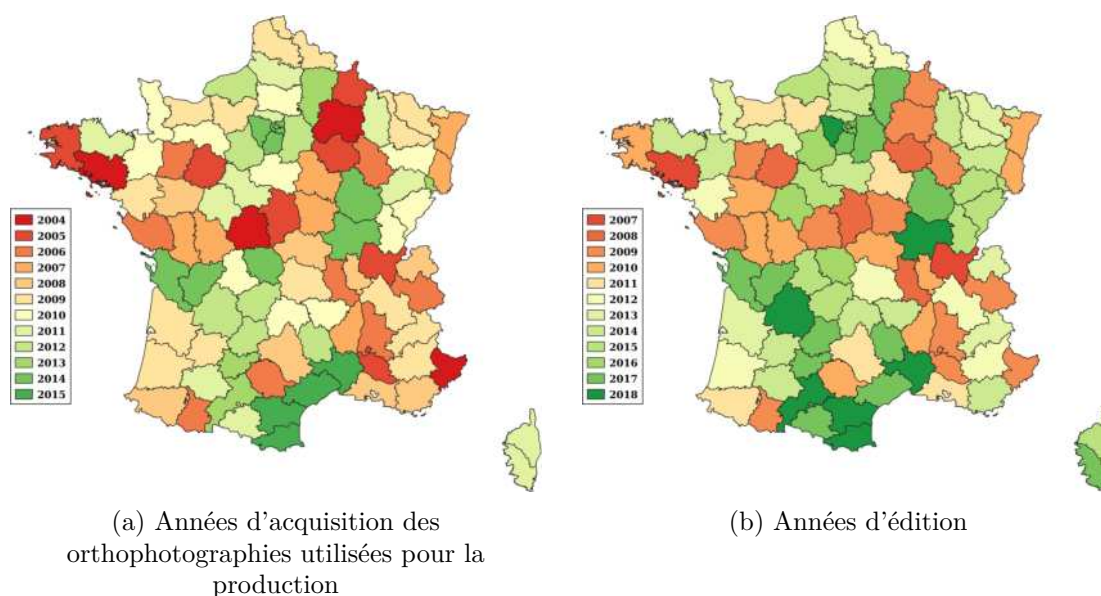


FIGURE 1.4 – Années d'acquisition (a) et de diffusion (b) de la BD Forêt V2 par département métropolitain.

La surface nationale des peupleraies issue de la BD Forêt a pu être déduite en 2018 suite

à la couverture de l'ensemble du territoire. Elle est égale à 206 751 ha. Outre l'incertitude liée au processus de calcul, cette valeur n'est pas représentative de l'état des surfaces en 2018 puisque basée sur des couvertures départementales datant de 2007 à 2018 (Figure 1.4a). Bien que le délai de production de la BD Forêt soit adapté à la grande majorité des formations forestières ayant des cycles sylvicoles de plus de 40 ans, il n'est pas suffisant pour le suivi des peupleraies en raison de leur croissance rapide et leur cycle de rotation court (15 à 20 ans).

L'inventaire forestier :

L'inventaire forestier, également produit par l'IGN, est une évaluation annuelle permanente des ressources forestières nationales. C'est une méthode statistique établie selon une maille carrée de 1 km de côté permettant de construire dix sous-ensembles d'échantillons annuels différents. Cette grille décennale est séparée en deux cycles quinquennaux de manière juxtaposée permettant de revisiter les points du premier cycle lors du deuxième passage, avec de manière générale, un point par maille.

L'estimation est établie en deux phases : la première consiste en une photo-interprétation de 100 000 points par an, complétée par une seconde phase de revisite annuelle d'un sous échantillon d'environ 14 000 points comportant des points nouveaux et des points revisités pour la seconde fois (cinq ans après la première visite) (IGN, 2018). Afin d'avoir des résultats plus précis et fondés sur plusieurs échantillons annuels, les estimations sont établies pour l'année médiane du cycle quinquennal avec une incertitude de l'ordre de 100 000 ha (IGN, 2019).

Bien que la méthode de l'inventaire soit ajustée pour mieux s'adapter aux peupleraies compte tenu de leur répartition territoriale inégale, en échantillonnant quatre points par maille dans les zones populicoles au lieu d'un seul, la surface estimée demeure peu précise en raison d'une marge d'erreur importante. Pour le cycle quinquennal 2015-2019, la surface nationale des peupleraies est évaluée à 163 000 ha \pm 19 000.

Le cadastre :

La surface nationale des peupleraies peut également être déduite à partir des déclarations cadastrales qui répertorient une classe de nature de culture correspondant aux peupleraies : Bois de Peuplier (BP). Toutefois, cette information n'est mise à jour que sur la base des déclarations transmises par les propriétaires.

Si la déclaration de la vocation des cultures est obligatoire, les données cadastrales ne sont pas régulièrement actualisées et les changements de nature de culture ne sont pas systématiquement signalés. Un extrait du plan cadastral sur le département de la Haute-Garonne superposé à l'orthophotographie de la même année est présenté dans la Figure 1.5. Il fait appa-

raître des déclarations en BP (polygones rouges) correspondant à des vergers ou des champs agricoles mais aussi des peupleraies non déclarées en BP. La surface cadastrale nationale déclarée en BP était de 233 985 ha en 2016.

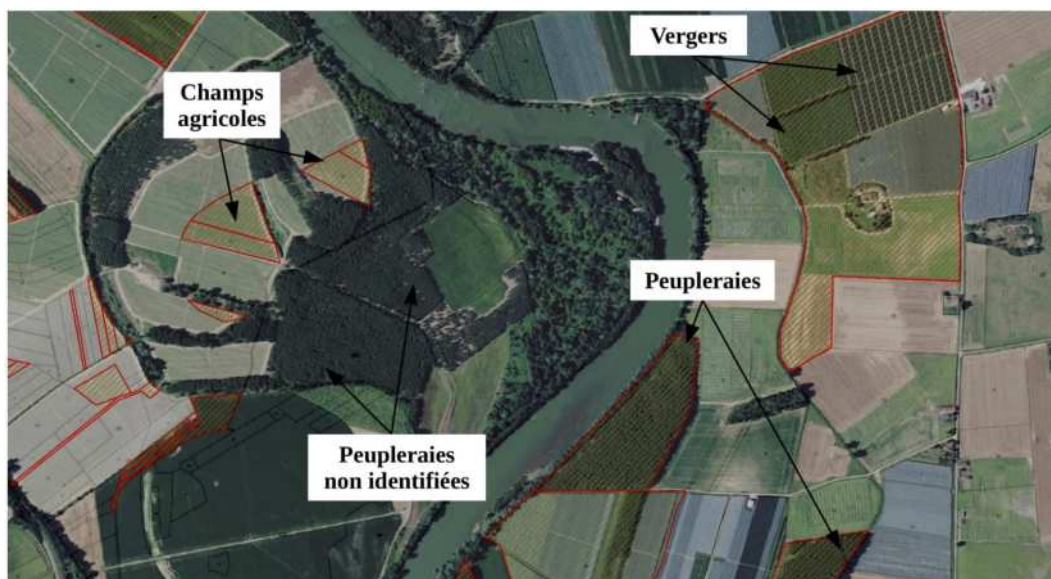


FIGURE 1.5 – Extrait du plan cadastral sur le département de la Haute-Garonne issu du cadastre de 2016 superposé sur l'orthophotographie de la même année. Les polygones rouges indiquent les déclarations de bois de peuplier (BP).

Les surfaces des peupleraies fournies par ces trois sources sont peu convergentes et difficilement comparables en raison de la divergence entre les méthodes de calcul et la disparité entre les années de mise à disposition de l'information. La rotation rapide des peupliers ainsi que la possibilité de changement de vocation des parcelles en fin du cycle rend l'évaluation des surfaces encore plus difficile.

La filière a recours à d'autres sources telles que les statistiques de vente de plants pour estimer la surface des plantations renouvelées et boisées, ou encore les enquêtes annuelles de branches (EAB) afin d'évaluer les volumes de bois récoltés et en déduire ainsi les surfaces exploitées. Ces données sont collectées par département et sont ensuite agrégées par région et au niveau national. Bien que jugées fiables par la filière, ces statistiques ne concernent que les changements (nouvelles plantations ou coupes) et ne peuvent être obtenues que de manière rétrospective. L'autre limite réside dans le fait que ces données ne sont pas spatialisées et qu'elles sont fournies de manière globale au niveau départemental.

Face à la difficulté de connaître et de suivre précisément la ressource populicole avec les sources de données existantes et afin de mieux sécuriser et développer la filière peuplier, la production de données cartographiques fiables et régulières s'avère cruciale.

1.2 La télédétection pour le suivi des milieux forestiers

La télédétection a depuis toujours été considérée comme un outil incontournable pour le suivi des milieux forestiers (Boyd, 2005). Au cours des quatre dernières décennies, les progrès significatifs associés aux avancées dans le développement des capteurs et la croissance exponentielle des techniques et outils d'extraction et de traitement de l'information ont permis d'aborder la question de la classification des espèces d'arbres sous différentes perspectives (Wulder and Franklin, 2003).

Cette section se focalise sur l'utilisation de la télédétection satellitaire pour l'observation de la surface de la Terre. Nous introduisons d'abord le principe de base de la télédétection en mettant l'accent sur la télédétection spatiale. Nous retraçons ensuite l'évolution des caractéristiques des capteurs optiques, et dans la dernière partie, nous présentons une synthèse des travaux pour le suivi des milieux forestiers, et notamment des plantations.

1.2.1 Généralités sur la télédétection spatiale

Au sens général du terme, la télédétection désigne l'ensemble des techniques d'observation de cibles et de phénomènes à la surface de la Terre ou d'autres planètes par le biais de capteurs situés à distance. Ces capteurs peuvent être embarqués à bord de satellites, d'avions, de drones télécommandés ou programmés ainsi que d'autres plateformes mobiles telles que les ballons et les bateaux.

Bien que son origine soit souvent confondue avec celle de la photographie aérienne, la télédétection a été réservée durant près d'un siècle à des besoins de reconnaissance militaire mobilisant des capteurs installés à bord d'avions.

Elle a véritablement pris son essor dans les années 1970 avec le déploiement des satellites d'observation de la Terre et le développement des capteurs numériques. Les applications civiles se sont dès lors multipliées dans les domaines terrestre, maritime et atmosphérique.

Principe :

Tous les systèmes de télédétection reposent sur le même principe de mesure de l'émission et/ou de la réflexion du rayonnement électromagnétique (REM). On distingue ainsi deux types de capteurs : les capteurs actifs et les capteurs passifs (Figure 1.6).

Dans le cas des capteurs passifs, le REM émis par le soleil est envoyé vers la surface. Les cibles illuminées absorbent une partie de cette énergie, une autre est transmise au milieu et le reste est réfléchi. Le capteur enregistre la partie réfléchi dans sa direction. Ils sont ainsi dépendants du soleil en tant que source d'énergie et le REM réfléchi ne peut donc être mesuré que de jour. Les systèmes actifs de télédétection produisent leur propre source d'énergie pour

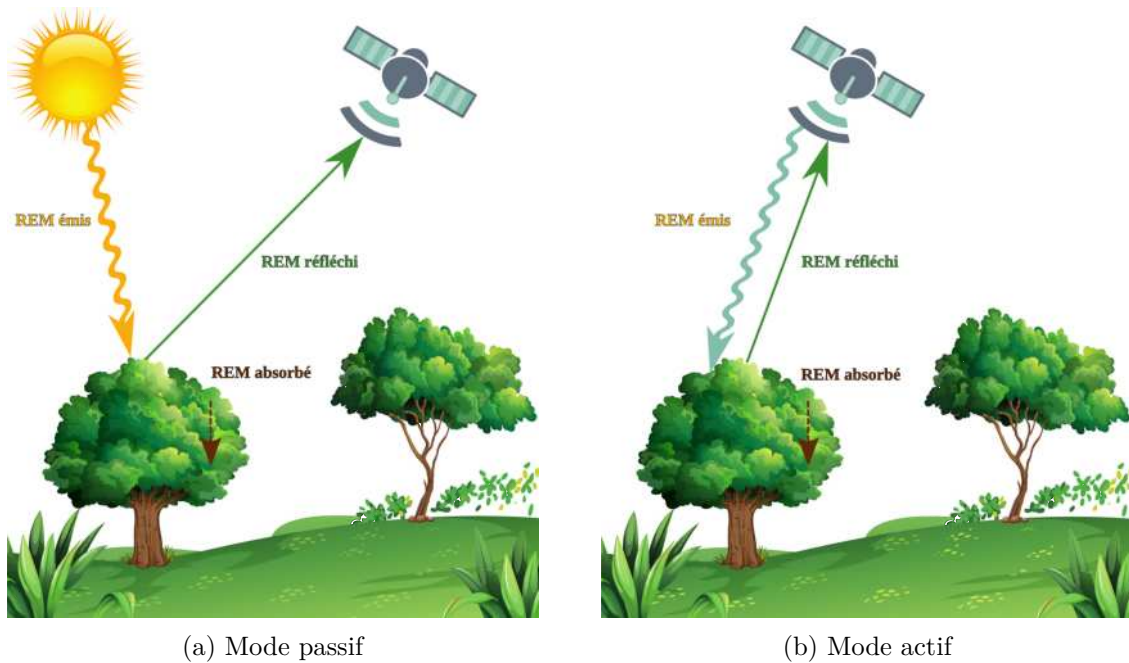


FIGURE 1.6 – Modes d'acquisition des systèmes de télédétection.

illuminer les cibles en surface. Une partie du REM émis est absorbée par la cible et l'autre est réfléchi pour être mesurée par le capteur. L'acquisition avec les capteurs actifs est donc possible de jour comme de nuit.

Avant d'atteindre les cibles, le REM émis par la source est intercepté par les particules atmosphériques. En fonction de sa longueur d'onde (λ), l'énergie incidente peut traverser ou être diffusée ou absorbée par l'atmosphère (Figure 1.7).

Les instruments de télédétection opèrent dans quasiment tout le spectre électromagnétique, mais principalement entre les gammes du visible et des micro-ondes (Figure 1.7). La télédétection optique utilise des capteurs qui fonctionnent généralement dans les longueurs d'ondes allant du domaine du visible ($\lambda = 400 \text{ nm}$) jusqu'au début de l'infrarouge lointain ($\lambda = 15 \mu\text{m}$). Dans le domaine optique, les longueurs d'ondes sont plus ou moins affectées par les phénomènes de diffusion et d'absorption atmosphériques limitant l'acquisition sous des conditions défavorables (ex. présence d'une forte couverture nuageuse). Ces perturbations sont variables dans le temps et dépendent de l'épaisseur atmosphérique ainsi que de sa concentration en aérosols. Les *Light Detection And Ranging* (LIDAR) et les caméras multispectrales sont des exemples de capteurs optiques.

Les capteurs de télédétection Radio Detection And Ranging (RADAR) utilisent des ondes plus longues dans la gamme des micro-ondes ou hyperfréquences. Les RADAR courants fonctionnent avec des ondes centimétriques ayant des longueurs d'ondes de 1 cm à 1 m. Ces ondes plus longues passent au travers de la couche nuageuse (atmosphère transparente) puisqu'elles ne sont pas sensibles à la diffusion atmosphérique qui affecte les ondes plus courtes. Cette

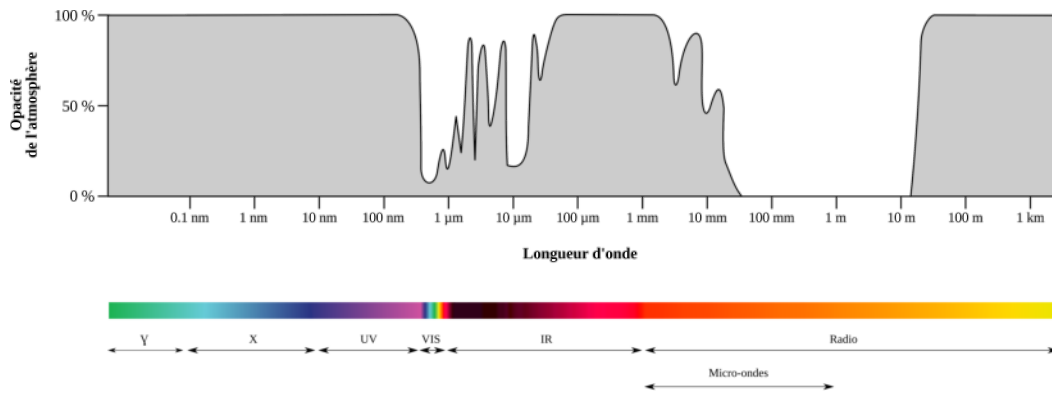


FIGURE 1.7 – Opacité atmosphérique (absorption gazeuse) en fonction de la longueur d’onde. La courbe ne tient pas compte des atténuations potentielles.

propriété permet la détection dans quasiment toutes les conditions atmosphériques. Il s’agit par exemple des satellites Radar à Synthèse d’Ouverture (RSO) ou des altimètres RADAR.

Caractéristiques de résolution :

Les données de télédétection ont des caractéristiques de résolution et de géométrie qui sont fonction du système d’acquisition. Ainsi, suivant le mode optique ou radar, ces caractéristiques ne sont pas définies de la même manière.

La suite de ce chapitre sera plus spécifiquement consacrée à l’imagerie satellitaire optique utilisée dans le cadre de ces travaux de thèse.

De manière générale, les images de télédétection optique sont caractérisées par quatre principales résolutions :

- **la résolution spatiale** : elle définit la taille du plus petit objet qu’il est possible de détecter et traduit ainsi la finesse du détail observable (Figure 1.8). En réalité, cette capacité à restituer les détails dépend de la fonction de transfert de modulation (FTM) instrumentale et de l’échantillonnage spatial, en plus du bruit (Lier et al., 2008). L’acceptation plus commune et simplifiée de cette notion est de considérer la longueur du côté d’un pixel au sol. La résolution spatiale est une caractéristique intrinsèque du capteur. Artificiellement, il est possible de procéder à un ré-échantillonnage en divisant ou en agrégeant des pixels voisins par exemple, mais la résolution native ne change pas.
- **la résolution spectrale** : elle reflète la capacité d’un capteur à distinguer des REMs de longueurs d’onde différentes. Le capteur agit comme un prisme qui, en fonction de sa résolution spectrale, va décomposer le REM réfléchi par les cibles et l’enregistrer dans une gamme de longueurs d’onde successives représentées par différentes bandes spectrales (Figure 1.9). La résolution spectrale est définie par la plus petite largeur de bande enregistrée, mais par abus de langage, elle est donnée par le nombre de bandes

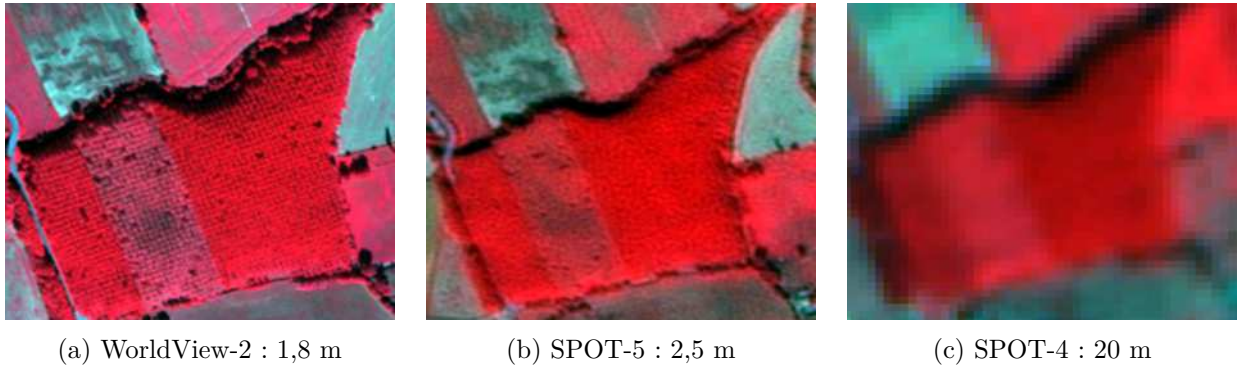


FIGURE 1.8 – Illustration de la notion de résolution spatiale sur des extraits d’images satellitaires en composition colorée infrarouge couleur d’une peupleraie dans les Coteaux de Gascogne.

spectrales. Plus la résolution spectrale est fine, plus la largeur des bandes est étroite. On parle ainsi de capteurs multispectraux qui possèdent trois à dix bandes larges et discrètes et de capteurs hyperspectraux avec plus de 100 bandes étroites et continues. La discrimination entre les cibles est donc possible par l’analyse des réponses spectrales dans les différentes longueurs d’onde, désignées comme la *signature spectrale*, qui est spécifique à chaque objet. Ce point sera abordé plus en détails dans la Section 1.2.1. Par exemple, pour des classes génériques telles que l’eau, la végétation et le sol, trois bandes larges du visible et une bande dans le domaine du proche infrarouge permettent sans difficulté de les distinguer (Figure 1.9). En revanche, pour la séparation de classes d’essences d’arbre, il est nécessaire de disposer de plus de bandes étroites.

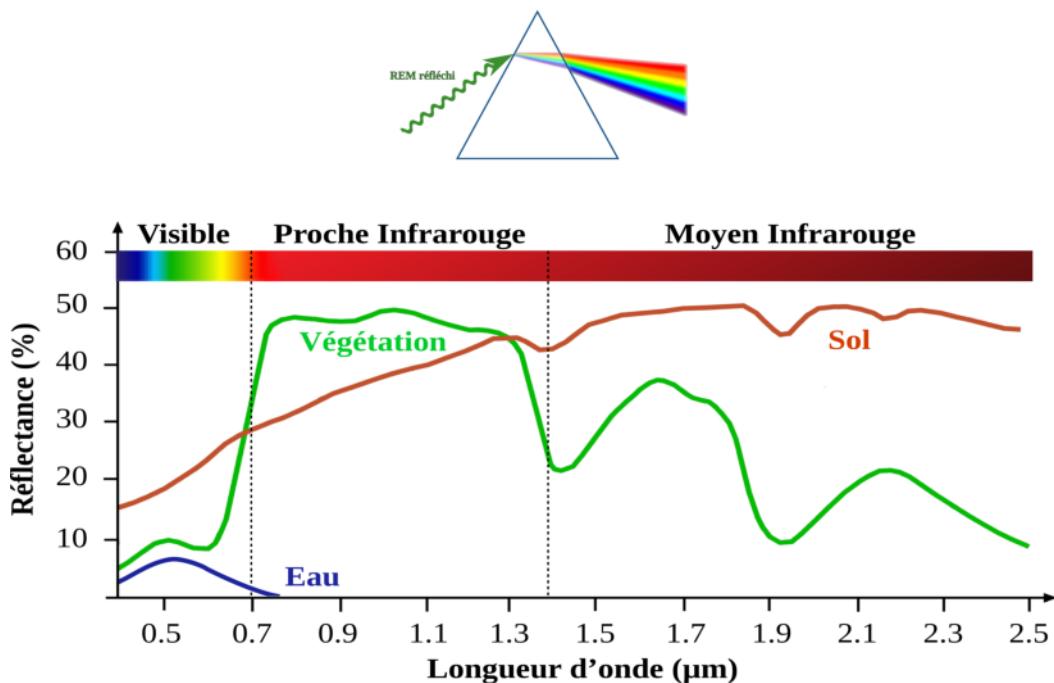


FIGURE 1.9 – Exemple de signatures spectrales de trois surfaces naturelles dans les domaines du visible, du proche infrarouge et du moyen infrarouge.

- **la résolution temporelle** : elle correspond au plus petit intervalle de temps entre deux acquisitions. On parle ainsi de séries temporelles d’images satellitaires lorsque la même scène est observée à différentes dates. La notion de résolution temporelle est importante en télédétection satellitaire car elle permet non seulement d’assurer un suivi régulier de l’évolution des caractéristiques spectrales dans le temps aux échelles intra et inter-annuelles mais aussi de détecter des changements. Un exemple de vues satellitaires d’une peupleraie entre août et octobre 2019 est présenté dans la figure 1.10. Elle montre une peupleraie mature le 24/08/2019 qui a commencé à être coupée le 13/09/2019. Au moment du passage du satellite, le 08/10/2019, toute la peupleraie était entièrement exploitée.



FIGURE 1.10 – Illustration de l’évolution d’une peupleraie entre août et octobre 2019 avec des images Sentinel-2 dans le département du Tarn-Et-Garonne.

En plus de la signature spectrale, l’information temporelle est parfois indispensable pour séparer les différents types de végétation. En effet, la succession des événements clés du cycle végétal, appelée phénologie, est différente en fonction de l’espèce. Le comportement temporel peut donc être un facteur déterminant pour une meilleure discrimination entre espèces (Fassnacht et al., 2016).

- **la résolution radiométrique** : elle correspond à la plus petite différence d’intensité du signal enregistré. Plus la résolution radiométrique d’un capteur est fine, plus il est sensible aux petites variations d’intensité de l’énergie reçue. La gamme des longueurs d’onde à laquelle le capteur est sensible est appelée plage dynamique. La résolution radiométrique dépend du nombre de bits enregistrés par pixel qui est utilisé pour encoder les valeurs au format binaire. Plus ce nombre est grand, plus la plage dynamique est importante. Par exemple, un capteur qui enregistre les données en 16 bits aura une plage dynamique de $2^{16} = 65\,536$ valeurs, et donc une résolution radiométrique plus fine qu’un capteur utilisant 8 bits avec $2^8 = 256$ valeurs d’intensité.

Propriétés optiques :

Quand le REM émis par le soleil atteint la surface illuminée, chaque cible observée va adopter un comportement spectral spécifique à chaque longueur d'onde. L'ensemble des réponses spectrales à différentes longueurs d'onde constitue ce que l'on appelle la signature spectrale. Outre la longueur d'onde λ et les conditions d'acquisition, le comportement spectral de chaque objet est régi par ses propriétés optiques, qui caractérisent sa capacité à absorber, transmettre ou réfléchir le REM incident. La suite de cette section sera uniquement consacrée à l'analyse des propriétés optiques de la végétation.

Dans le cas de la végétation, le feuillage constitue le principal élément d'interaction avec le REM incident et ce sont donc les propriétés optiques des feuilles qui vont déterminer le comportement spectral. Les facteurs d'absorption, de transmission et de réflexion sont dépendants des caractéristiques biochimiques des feuilles qui seront plus ou moins déterminantes en fonction de la région spectrale considérée.

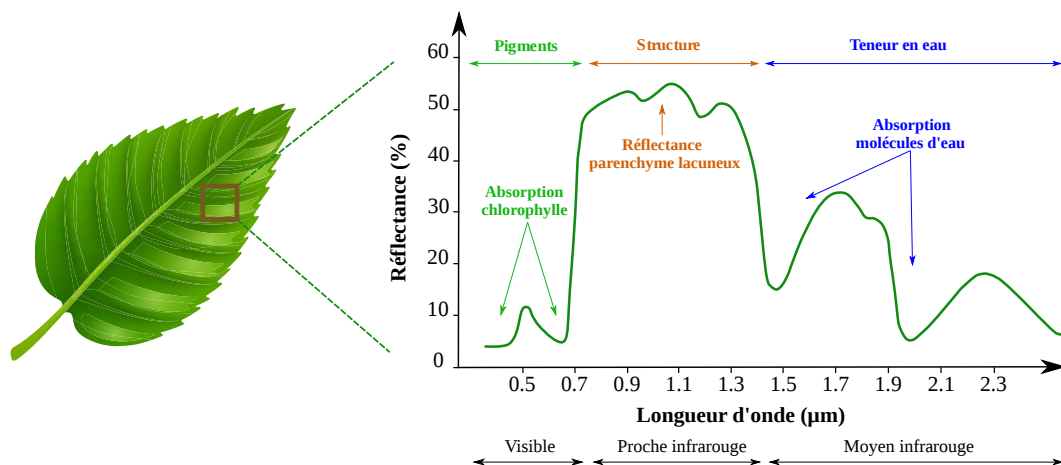


FIGURE 1.11 – Interaction entre le rayonnement électromagnétique et la feuille selon le domaine spectral. Figure établie d'après Girard et al. (1999).

On distingue trois domaines spectraux :

- **le visible (0,4 - 0,7 μm)** : dans ce domaine, ce sont les pigments foliaires qui influencent le comportement spectral de la végétation. Il s'agit des anthocyanes (pigments rouges à bleus), des carotènes (pigments orangés) et plus particulièrement des chlorophylles (pigments verts) qui présentent deux bandes d'absorption dans le bleu et dans le rouge (Devineau, 1990; Jacquemoud et al., 2009) (Figure 1.11). La présence des pigments foliaires est donc responsable de la forte absorption du rayonnement visible, avec un maximum de réflectance dans le vert (faible absorption). Si les pigments chlorophylliens absorbent la majeure partie du spectre visible, ils sont totalement transparents dans les rayonnements à longueur d'onde plus importante.
- **le proche infrarouge (0,8 - 1,1 μm)** : dans le proche infra-rouge, les ondes électro-

magnétiques sont très faiblement absorbées ($< 10\%$) et sont essentiellement réfléchies ou transmises. Le rayonnement infrarouge traverse ainsi les couches supérieures de la feuille composées notamment par la cuticule, l'épiderme et le parenchyme palissadique renfermant la majorité des chloroplastes qui sont transparents à ces longueurs d'onde, pour atteindre enfin le parenchyme lacuneux (pauvre en chloroplastes). Plus la structure du parenchyme lacuneux est développée, plus le rayonnement est réfléchi à ce niveau. L'infrarouge est particulièrement intéressant pour différencier les feuillus des conifères qui présentent généralement un parenchyme lacuneux plus développé, se traduisant par une réflectance supérieure d'au moins 10% à celle des résineux (Girard et al., 1999).

- **le moyen infrarouge (1,1 - 2,5 μm)** : dans le moyen infrarouge, c'est principalement la teneur en eau des feuilles qui gouverne. Plus la teneur en eau est forte, plus la réflectance de la végétation est faible, particulièrement aux longueurs d'onde 1,4 μm et 1,9 μm qui correspondent aux bandes d'absorption de l'eau (Figure 1.11). Une absorption par la matière sèche (cellulose, lignine) a également lieu dans le moyen infrarouge mais dans une moindre mesure et seulement lorsque la feuille se dessèche (Clark and Roberts, 2012).

1.2.2 Contexte actuel : évolution des caractéristiques des capteurs de télé-détection satellitaire

Depuis le lancement des capteurs d'observation de la Terre dans les années 1970, les satellites imageurs opérationnels ont supplanté les approches traditionnelles d'estimation de la couverture forestière (Fuller, 2006), et plus généralement, de suivi de la surface de la Terre.

La première initiative destinée à des applications civiles a été le programme Landsat lancé par la *National Aeronautics and Space Administration* (NASA). Il a été consacré à la surveillance régulière des ressources terrestres et à l'estimation du rendement des récoltes céréalières aux États Unis et en Russie afin d'anticiper l'évolution des cours boursiers. Le premier satellite Landsat-1 était équipé de l'instrument *Multispectral Scanner System* (MSS) lui permettant d'acquérir des scènes de plus de 170 km² dans quatre bandes spectrales et avec une résolution spatiale décimétrique (80 m). Les scènes étaient observées avec une répétitivité de 18 jours.

Vers la fin des années 1990, et avec la sensibilisation aux enjeux mondiaux liés au changement climatique et la nécessité croissante de suivre de manière régulière et continue tout le système climatique de la Terre, d'autres programmes spatiaux ont vu le jour. Il s'agit notamment des missions Terra et Aqua lancées dans le cadre du programme *Earth Observing System* (EOS) de la NASA. Les deux satellites sont dotés des capteurs Moderate Resolution Imaging Spectroradiometer (MODIS). Ils observent toute la surface de la Terre à raison d'une

image par jour dans 36 bandes spectrales à des résolutions spatiales allant de 250 m à 1 km permettant ainsi de capturer des dynamiques spatio-temporelles de manière globale et régulière. Cependant, les résolutions spectrales et temporelle exceptionnelles des instruments MODIS sont obtenues au détriment d'une résolution spatiale kilométrique, ce qui limite leur application à des échelles locales requérant un niveau de détail élevé.

Par ailleurs, des satellites de haute à très haute résolution spatiale à vocation commerciale se sont développés. Il y a d'abord eu la première génération de satellites français Satellites Pour l'Observation de la Terre (SPOT)¹ avec une résolution spatiale de 10 à 20 m dans quatre bandes spectrales. Ils ont été suivis par les satellites américains Ikonos² et QuickBird³ avec des résolutions spatiales allant respectivement jusqu'à 1 m et 65 cm, puis par WorldView en 2007 (jusqu'à 50 cm de résolution spatiale et une fauchée large de 17 km) et Pléiades en 2009 (jusqu'à 70 cm de résolution spatiale et une fauchée large de 20 km). Cependant, le coût élevé ainsi que la faible couverture de ces capteurs ne permettaient pas d'assurer un suivi sur de grandes étendues.

Le développement de technologies pour les capteurs et les instruments de télédétection, parallèlement avec l'augmentation des besoins et des applications spécifiques, ont contribué à la conception de missions d'observation de la Terre plus ciblées et à haute performance ainsi qu'à l'expansion et l'amélioration des programmes spatiaux. La série Landsat, notamment, compte aujourd'hui huit satellites dont deux actuellement opérationnels et un neuvième qui devrait être mis en orbite en septembre 2021. Le capteur *Operational Land Imager* (OLI) embarqué à bord du Landsat-8 permet d'acquérir des images de 15 à 30 m de résolution spatiale selon neuf bandes spectrales du visible au moyen infrarouge. Son deuxième capteur *Thermal Infrared Sensor* (TIRS) mesure plutôt la température de surface de la Terre dans deux bandes en infrarouge thermique à 100 m de résolution spatiale. Ces images sont acquises avec un temps de revisite de 16 jours.

Grâce à des investissements publics et privés conséquents dans la diffusion gratuite des images de télédétection, l'utilisation des données satellitaires s'est démocratisée. Depuis 2013, les archives et les nouvelles acquisitions Landsat sont mises à disposition gratuitement par l'*United States Geological Survey* (USGS). Plus récemment, l'agence spatiale européenne ou *European Space Agency* (ESA) a lancé le programme Copernicus⁴ d'observation et de surveillance de la Terre. Opérationnel depuis 2014, Copernicus est considéré comme le programme le plus ambitieux à ce jour. Il consiste en une constellation de six satellites « les Sentinel » destinée à fournir gratuitement les données et les services nécessaires au suivi opérationnel

1. SPOT-1 lancé en 1986, fauchée de 60 km x 60 km

2. Ikonos-2 lancé en 1999, fauchée de 11 km x 11 km

3. QuickBird I lancé en 2000, fauchée de 18 km x 18 km

4. L'ancien programme Global Monitoring for Environment and Security (GMES) de surveillance de la Terre

des effets du changement climatique.

En outre, le satellite hyperspectral italien de la mission *PRecursore IperSpettrale della Missione Applicativa* (PRISMA) est mis en orbite depuis 2019. Il s'agit de la première mission satellitaire hyperspectrale en Europe qui permet d'acquérir des données hyperspectrales sur 239 bandes fines de moins de 12 nm entre 0,4 μm à 2,5 μm . Les images sont à 30 m de résolution spatiale et acquises avec une répétitivité de 29 jours. Les premiers travaux réalisés avec les données PRISMA montrent le potentiel de la mission pour la discrimination des principaux types forestiers (Vangi et al., 2021), mais d'autres applications suivront avec l'augmentation de la disponibilité des acquisitions. D'autre part, la mission allemande *Environmental Mapping and Analysis Program* (EnMAP) prévue pour 2022, embarquera la caméra *Hyperspectral Imager* (HSI) qui fonctionnera dans le spectre du visible au moyen infrarouge afin de fournir des observations hyperspectrales inédites à l'échelle mondiale. Les images seront acquises dans 232 bandes spectrales à une résolution spatiale de 30 m et avec une répétitivité de quatre jours. Une synthèse plus exhaustive sur les principales missions satellitaires actuelles et celles prévues à l'horizon 2030 est fournie dans Ustin and Middleton (2021).

La dernière décennie a été marquée par une nouvelle ère technologique en télédétection. Pour le suivi des milieux forestiers en particulier, l'amélioration de l'information temporelle, spectrale ainsi que la résolution spatiale offerte par la nouvelle génération de satellites a favorisé un suivi plus fin sur de grandes étendues. Le nombre d'études sur la cartographie d'essences forestières a également suivi l'augmentation de la disponibilité des données à haute résolution, avec une tendance quasi exponentielle depuis 2010 (Fassnacht et al., 2016).

Ces nouvelles opportunités s'accompagnent de défis principalement liés aux problèmes de gestion des séries temporelles denses à haute dimension spectrale, à la complexité de la surface observée et à la variabilité de son comportement spectral dans l'espace et dans le temps, mais également à des contraintes techniques et méthodologiques lorsqu'il s'agit de suivre de grandes étendues à des échelles nationales ou continentales.

La section suivante présente les principales approches méthodologiques adoptées en télédétection pour le suivi des essences forestières.

1.2.3 Télédétection forestière : un état de l'art

À chaque application, son contexte et ses contraintes

La littérature propose un grand nombre d'études sur le suivi des essences forestières par télédétection, avec différents **types de capteurs**, en privilégiant une ou plusieurs **résolutions** afin de caractériser la **composition** de la ressource forestière à différentes **échelles spatiales**. Ces études abordent le sujet à travers une multitude d'applications thématiques qui déterminent le choix des caractéristiques relatives à ces quatre volets résumés de manière

synthétique dans la carte heuristique de la Figure 1.12. Ces applications comprennent, entre autres, la cartographie de la composition des forêts en essences (Ørka et al., 2013), le suivi de la biodiversité dans les forêts anciennes (Spracklen and Spracklen, 2019), l'estimation de la biomasse et de la productivité des forêts (Lefsky et al., 2005) ou encore l'analyse de la réponse des forêts au stress ou au changement climatique (Lausch et al., 2016; Asner and Alencar, 2010).

Dans cette partie, il ne s'agit pas de fournir une vue d'ensemble exhaustive de ces diverses applications et leurs approches sous-jacentes, mais de présenter de manière synthétique les principales orientations adoptées en télédétection pour la cartographie des essences forestières, suivies de quelques exemples de travaux.

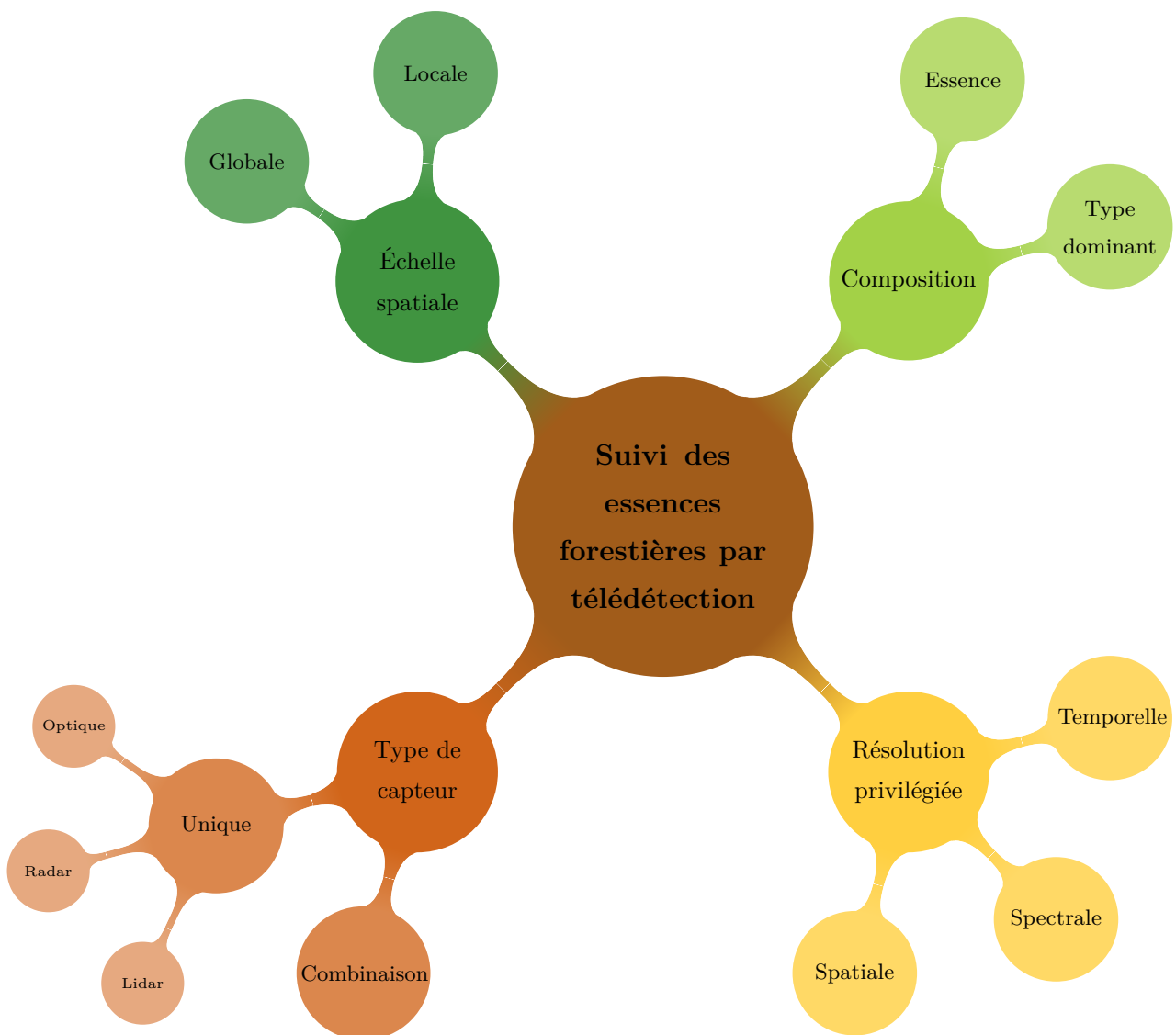


FIGURE 1.12 – Carte heuristique des principales composantes considérées dans la littérature pour le suivi des essences forestières.

À une échelle donnée et en fonction de la technologie du capteur utilisé, différentes informations peuvent être extraites pour l'identification des essences.

Les capteurs optiques multispectraux ou hyperspectraux sont généralement exploités grâce à leur richesse spectrale permettant de séparer différentes essences en mesurant la réponse réfléchie par la canopée dans les différentes longueurs d'onde possibles du capteur. Cependant, cette capacité de séparation diffère d'un capteur à l'autre selon les caractéristiques de résolution (radiométrique, spectrale, spatiale et temporelle). L'information spatiale liée à la texture, définie comme la variation des niveaux de gris dans une image, est souvent dérivée de capteurs optiques à très haute résolution spatiale permettant de capturer des motifs fins spécifiques à chaque essence, mais aussi à partir de données RADAR de par leur sensibilité à la rugosité des surfaces observées. Les données issues des capteurs LIDAR et RADAR permettent de déterminer des caractéristiques structurales liées notamment à l'arrangement des arbres dans un peuplement ou à la densité des branches qui définit la structure verticale. Des caractéristiques temporelles sont généralement extraites à partir de séries temporelles d'images optiques ou RADAR et permettent de dériver des indicateurs phénologiques généralement pertinents pour séparer les essences observées. D'autres caractéristiques telles que l'information contextuelle ou les traits physiologiques peuvent également être extraites.

À l'échelle globale, ce sont les capteurs à faible résolution spatiale et à large couverture qui sont privilégiés. Par exemple, [Friedl et al. \(2002\)](#) ont utilisé des images MODIS pour cartographier 17 classes d'occupation des sols dont cinq principaux types de forêt à l'échelle mondiale avec une résolution spatiale de 1 km. Malgré leur relative richesse spectrale, leur couverture globale et leur fréquence d'acquisition quotidienne qui augmentent les chances d'obtenir des images sans nuages ([Hansen et al., 2008](#); [Dong et al., 2012](#)), la résolution spatiale grossière des capteurs MODIS rend difficile la caractérisation au niveau de l'essence en raison du mélange entre plusieurs espèces au sein du même pixel ([Dong et al., 2012](#); [Cano et al., 2017](#)). Ceci les rend moins pertinents pour des analyses fines à des échelles locales ([Hansen et al., 2008](#); [Walker et al., 2012](#)).

Depuis plus de 40 ans, beaucoup d'attention a été accordée aux capteurs multispectraux Landsat *Thematic Mapper* (TM) et *Enhanced Thematic Mapper Plus* (ETM+) ([Banskota et al., 2014](#)) pour la cartographie plus fine des principaux types forestiers ([Woodcock et al., 1994](#); [Kati Salovaara et al., 2005](#); [Banskota et al., 2014](#)), le suivi global de la déforestation ([Lehmann et al., 2013](#); [Kim et al., 2014](#)), la prédiction de variables biophysiques à des échelles locales ([Phua and Saito, 2003](#); [Cohen et al., 2003](#); [Main-Knorn et al., 2013](#)) et un peu moins pour le suivi phénologique des forêts ([Melaas et al., 2013](#)). La présence de plusieurs années d'archive Landsat et l'amélioration de la résolution spatiale comparée à MODIS a permis d'assurer des suivis à différentes échelles sur le long terme. Cependant la densité temporelle limitée des acquisitions (16 jours) augmente la probabilité d'obtenir des données nuageuses et limite ainsi les applications de suivi phénologique des forêts et de la détection de changements

intra-annuels. Les approches visant à combiner les images Landsat avec MODIS constituent des pistes alternatives pour contourner leur résolution temporelle limitée (Gao et al., 2006; Hansen et al., 2008; Walker et al., 2012).

L'arrivée des satellites Sentinel-2 avec des résolutions spatiale (10 à 20 m), spectrale (13 bandes spectrales) et temporelle (5 jours) a fourni de nouvelles opportunités pour la classification fine au niveau de l'espèce. Les travaux avec les données Sentinel-2 de Immitzer et al. (2016); Persson et al. (2018); Bolyn et al. (2018); Immitzer et al. (2019); Karasiak et al. (2020); Axelsson et al. (2021) s'inscrivent dans la lignée des travaux antérieurs basés sur l'information temporelle (Hill et al., 2010; Stoffels et al., 2015).

Les capteurs RADAR satellitaires ont été moins impliqués dans le suivi précis des essences forestières que les capteurs optiques. Les acquisitions à résolution spatiale modérée ont historiquement été mobilisées pour l'identification globale du couvert forestier et le suivi de la déforestation (Shimada et al., 2014; Zhang et al., 2012). De nouvelles perspectives sur l'utilisation du RADAR pour une caractérisation plus fine des essences commencent à émerger avec l'arrivée des séries temporelles Sentinel-1 à haute résolution spatiale (Dostálová et al., 2018a; Rüetschi et al., 2018; Frison et al., 2018).

Souvent combinée à des données optiques hyperspectrales, l'information tridimensionnelle dérivée des capteurs LIDAR a fait l'objet de plusieurs travaux sur le cartographie d'essences forestières (Jones et al., 2010; Ghosh et al., 2014). Cependant leur coût d'acquisition élevé et leur couverture limitée ne permettent pas leur utilisation sur de grandes étendues.

L'information texturale a également été exploitée dans le cadre de travaux d'identification d'essences forestières. Elle a généralement été extraite à partir de données à très haute résolution spatiale pour une caractérisation à un niveau local (Kim et al., 2009; Mallinis et al., 2008). Elle est souvent combinée avec l'information spectrale pour une meilleure discrimination des espèces (Franklin et al., 2000; Mallinis et al., 2008).

Dans cette partie nous avons montré que pour la cartographie des essences forestières, différentes approches peuvent être adoptées en fonction de l'échelle et des caractéristiques du capteur considéré. Il s'agit seulement d'un bref aperçu des différents travaux. Un état de l'art plus détaillé est fourni dans la revue de littérature de Fassnacht et al. (2016).

Téledétection des plantations, pas d'exceptions

Bien qu'il n'y ait pas encore de réponses claires sur les spécifications requises, pour obtenir une caractérisation jusqu'au niveau de l'espèce, une information spectrale fine à des résolutions spatiale et temporelle suffisantes est nécessaire. Cela vaut également pour les plantations, qui sont généralement des peuplements homogènes composés majoritairement d'une seule espèce dominante. La cartographie des plantations forestières et de leurs dynamiques

spatio-temporelles a notamment pu tirer profit des récentes améliorations apportées aux capteurs spatiaux de télédétection. Plusieurs types de plantations ont fait l'objet de suivis par télédétection (Torbick et al., 2016). Il s'agit principalement de l'hévéa (ou arbre à caoutchouc) (Li and Fox, 2012; Ye et al., 2018; Xiao et al., 2019), du palmier à huile (Cheng et al., 2018; Puttinaovarat and Horkaew, 2019; Descals et al., 2019) de l'eucalyptus (le Maire et al., 2014; Forstmaier et al., 2020; Deng et al., 2020), et de l'acacia (Spracklen and Spracklen, 2021).

Pour identifier les plantations parmi les autres espèces forestières, les auteurs ont abordé la question de diverses manières en utilisant différentes sources de données et en privilégiant une ou plusieurs résolutions. Certains auteurs ont utilisé des images multispectrales mono-dates et à haute résolution spatiale pour discriminer des plantations monospécifiques de feuillus (Chardenon and Flouzat, 1981; Li and Fox, 2011; Eslami and Zahedi, 2011). Cependant, en raison de l'absence d'une information spectrale saisonnière, nécessaire pour distinguer les différents comportements phénologiques, les plantations ont souvent été confondues avec des espèces à feuilles persistantes ou avec des forêts naturelles.

D'autres auteurs ont plutôt privilégié l'information temporelle. Pour la cartographie des plantations d'hévéas à l'échelle de l'île de Hainan au Sud de la Chine, Han et al. (2018) ont utilisé trois années d'acquisitions Landsat multi-temporelles. Ils ont montré l'intérêt des indicateurs phénologiques basés sur la différence de réflectance dans le moyen infrarouge entre deux phases clés du cycle (foliation et défoliation) ce qui leur a permis de discriminer les plantations avec une précision globale de 93%.

Sur une plus large étendue en Chine, Deng et al. (2020) ont proposé une approche d'identification des plantations d'eucalyptus dans la région de Guangxi (superficie de 237 600 km²) en combinant les informations spatiale et temporelle dérivées des capteurs MODIS et Landsat ainsi que l'information spectrale issue de Sentinel-2. À une résolution spatiale de 30 m, les performances obtenues étaient limitées avec une mesure de rappel de 64% pour les eucalyptus, notamment en raison de la faible densité temporelle (celle de Landsat) dans des régions à forte couverture nuageuse. D'autres approches multi-temporelles ont également été explorées. Certains travaux ont analysé les variations saisonnières en calculant des indicateurs phénologiques Li and Fox (2012); Xiao et al. (2019); Poortinga et al. (2019) tandis que d'autres ont extrait des métriques saisonnières de texture Gao et al. (2015); Descals et al. (2019) ou des caractéristiques issues de séries temporelles telles que les « *shapelets* » Ye et al. (2018).

Télédétection des peupleraies depuis 1981

Bien que des efforts importants aient été consacrés à l'identification des peupliers naturels (Heyman et al., 2003; Viinikka et al., 2020; Kivinen et al., 2020), principalement des espèces forestières de peuplier tremble (Section 1.1.1), très peu de travaux ont porté sur la cartographie

des plantations de peuplier dans un contexte de forêt tempérée (Chardenon and Flouzat, 1981; Borry et al., 1993; Grignetti et al., 2009; Eslami and Zahedi, 2011).

Une des premières applications de la télédétection des peupleraies remonte à 1981 avec les travaux pionniers de Chardenon and Flouzat (1981) dans les vallées du Pô (en Italie) et de la Garonne (en France). Les auteurs ont utilisé les premières acquisitions Landsat-2 pour identifier des peupleraies de différentes classes d'âge et de divers cultivars. Avec des images mono-dates, la performance d'identification a atteint 81% pour les plantations matures alors qu'elle était inférieure à 50% pour les jeunes plantations en raison de la forte proportion de sol nu et de la résolution spatiale insuffisante (80 m). Ils ont souligné l'importance de l'information temporelle pour améliorer l'identification des peupleraies des autres peuplements forestiers mais aussi pour séparer les cultivars grâce au décalage phénologique entre les différents groupes (cultivars précoces ou tardifs).

Dans deux sites populicoles au Nord de l'Italie, Grignetti et al. (2009) ont exploité des images à très haute résolution spatiale Ikonos (1 m) et QuickBird (2,4 m) mono-dates acquises en juin 2000 pour identifier des peupleraies avec quatre classes d'âge. Ils ont travaillé à l'échelle des peuplements pour lesquels ils ont extrait des indices spectraux et des métriques texturales. Les résultats montrent des valeurs de précision globale comprises entre 74% et 82% avec des performances systématiquement plus faibles pour les jeunes plantations.

À notre connaissance, il n'y a pas eu, au cours de la dernière décennie, de travaux qui se sont focalisés spécifiquement sur l'identification des peupleraies par télédétection. Néanmoins, le peuplier a été classé comme une espèce parmi d'autres dans le cas de la cartographie de la composition en essences des forêts (Sheeren et al., 2016; Bolyn et al., 2020; Karasiak et al., 2021). Par exemple, dans Bolyn et al. (2020), les auteurs ont combiné des données optiques et RADAR pour la cartographie de neuf principaux types forestiers, dont le peuplier, en Belgique et dans le Nord de la France. Afin de s'affranchir de la couverture nuageuse, ils ont généré une seule image de synthèse à partir de toutes les acquisitions Sentinel-2 de l'année 2018 durant la saison de la végétation (mai à septembre). Ils ont cependant exploité l'information temporelle des données RADAR en calculant 12 synthèses mensuelles issues des séries temporelles Sentinel-1 de la même année. Ils ont obtenu des mesures de rappel pour la classe peuplier variant entre 57% et 78% dans les sites français et belge respectivement, une performance principalement pénalisée par la présence d'une classe de mélanges de feuillus souvent confondue avec les peupliers.

D'autres applications récentes ont porté sur des peupliers naturels, principalement dans un contexte de forêts boréales. Il s'agit par exemple des travaux de Viinikka et al. (2020) qui ont utilisé des données hyperspectrales HySpex (460 bandes de 455 à 2500 nm) à très haute résolution spatiale (0,5 et 1 m) pour la discrimination d'arbres individuels de peuplier tremble

en Finlande. Ils ont montré le potentiel de l'information spectrale en identifiant des régions spécifiques capables de distinguer, à l'échelle de l'arbre, les peupliers des autres essences. Bien qu'il s'agisse théoriquement de la même essence, ces approches concernent l'identification d'arbres individuels qui se distinguent des plantations tout au moins par leur physionomie et leur structure, qui sont fonction du cultivar planté, mais aussi par leur répartition et leur agencement dans l'espace. En effet, dans les peupleraies, les cultivars plantés ont généralement une certaine forme avec une croissance plutôt en longueur et une structure marquée par un tronc principal et très peu de branches basses, il s'agit d'arbres équidistants et du même âge. En revanche, les peupliers naturels sont moins agrégés dans l'espace et développent spontanément leurs ramifications et leurs pousses latérales. À ce stade, nous ignorons si ces différences structurelles et physionomiques entre les peupliers naturels et plantés seront marquées par un comportement spectral différent.

Dans cette partie nous avons présenté les principales approches entreprises pour l'identification d'essences forestières, soit en caractérisant la composition des forêts en termes des principaux types dominants, soit en cherchant à discriminer une seule espèce, tout en privilégiant une ou plusieurs caractéristiques de résolution.

Pour une caractérisation au niveau de l'espèce, et notamment pour un type de plantation donné, d'après les différents travaux, nous pouvons faire l'hypothèse que la combinaison d'une information multi-temporelle avec des résolutions spatiale et spectrale fines est nécessaire pour une meilleure séparation entre les espèces, et donc une meilleure discrimination.

Pour les peupliers en particulier, peu d'intérêt a été accordé à cette espèce dans les forêts naturelles et encore moins dans les plantations. La différence entre le comportement spectral et temporel des peupliers naturels et plantés n'a jamais été examinée et compte tenu des différences physionomiques et structurelles, il est difficile de savoir si une méthode développée pour identifier des peupliers naturels serait également adaptée aux peupleraies et inversement. Par ailleurs, il n'y a pas eu d'approches méthodologiques clairement définies et dédiées à la discrimination de cette essence. De plus, nous ne disposons pas de réponses claires sur l'intérêt d'incorporer ou non l'information temporelle pour mieux identifier les peupleraies.

En outre, la majorité des travaux précités ont été limités à des zones géographiquement restreintes avec peu de variabilité spatiale de l'essence observée ou peu de différences entre les conditions d'acquisition des images, voire aucune s'il s'agit d'une seule emprise ou d'images mono-dates, ce qui rend difficile l'évaluation des approches proposées sur de grandes étendues. En revanche, les quelques études menées à des échelles plus globales proposent des produits avec un niveau de détail grossier qui ne permet pas une identification jusqu'au niveau de l'espèce.

Bien que la production de cartes à l'échelle continentale ou mondiale soit facilitée par les

observations synoptiques et périodiques d'images de télédétection, il s'agit principalement de produits d'occupation des sols comprenant des classes de forêts généralistes sans caractérisation au niveau des espèces. Le niveau de détail thématique de ces produits reste insuffisant pour les gestionnaires forestiers et il n'est pas spécifiquement adapté à chaque application faisant de la classification des essences forestières sur de grandes étendues un des défis majeurs de ce domaine.

Dans la prochaine section, nous présentons les contraintes associées à la génération de produits cartographiques sur de grandes étendues. Nous identifions ensuite les principales approches adoptées en télédétection et, plus largement, en apprentissage automatique pour le passage à l'échelle.

1.3 La cartographie sur de grandes étendues par télédétection

1.3.1 Les contraintes pour le passage à l'échelle

Pour la génération de produits cartographiques sur de grandes étendues, il existe différentes contraintes à prendre en compte qui sont, de manière générale, soit associées aux satellites optiques imageurs, soit liées à la surface observée, notamment la végétation.

Pour les caractéristiques spécifiques aux capteurs, la cartographie sur de grandes étendues nécessite une couverture d'images suffisamment large (Wulder et al., 2012). Ainsi, la majorité des travaux à l'échelle globale se sont appuyés sur des données MODIS ou Landsat dont la largeur de fauchée est respectivement de 2330 et 185 km. La couverture globale doit également être considérée en parallèle avec la résolution spatiale mais aussi temporelle. Il s'agit d'un compromis de résolutions. En effet, la résolution spatiale est généralement inversement liée à la fauchée et à la répétitivité. Les données à résolution spatiale modérée telle que MODIS sont associées à une couverture globale et une très haute résolution temporelle. Pour les capteurs optiques, les observations répétées de la même scène augmentent les chances d'obtenir des données sans nuages, même dans les zones où la couverture nuageuse est permanente (Gómez et al., 2016). Cependant, avec une résolution spatiale grossière les dynamiques locales et détaillées sont difficiles à percevoir. À une résolution spatiale plus fine comme avec les données Landsat, la fauchée et surtout le cycle de revisite sont dégradés d'où le recours à des stratégies de fusion entre Landsat et MODIS pour favoriser les trois critères. Les données utilisées dans les initiatives de cartographie à grande échelle ne provenaient pas, jusqu'à récemment, de capteurs associant une large couverture avec une haute résolution spatiale et temporelle (Atzberger, 2013).

Afin de permettre des applications sur de grandes étendues, il est nécessaire de combiner une couverture globale avec des fréquences temporelles et des détails spatiaux suffisamment

élevés pour appréhender les dynamiques locales, trois caractéristiques améliorées avec les nouveaux satellites Sentinel-2.

D'autre part, pour un capteur donné, les acquisitions peuvent être hétérogènes pour diverses raisons telles qu'un échantillonnage temporel irrégulier d'un site à l'autre, une variation de l'éclairement ou la présence d'une couverture nuageuse de manière inégale entre les différents sites. Il est donc nécessaire d'envisager des approches adaptées et robustes aux variations des conditions d'acquisition.

Au-delà des contraintes associées aux capteurs, il existe des difficultés liées à la surface observée, et plus particulièrement à sa variabilité spatiale et temporelle.

Dans le cas particulier de la végétation, et pour une espèce donnée, la signature spectrale suit une dynamique saisonnière qui est fonction de son cycle phénologique (Reed et al., 1994). Cependant, en réponse aux conditions climatiques locales, la phénologie peut varier d'une région à l'autre. Le spectre observé est par conséquent différent pour une même espèce située à deux endroits différents. La variabilité spatiale de la phénologie est d'autant plus prononcée que l'étendue géographique est importante (Meroni et al., 2014). D'autres facteurs stationnels (topographie, type de sol, pratiques culturales, etc.) ou internes à l'espèce (état sanitaire, âge, cultivar) peuvent également influencer la réponse spectrale et induire une variabilité intra-espèces plus ou moins importante. À cela s'ajoute le fait que les acquisitions ne sont pas réalisées au même moment et avec la même qualité d'un endroit à l'autre, ce qui peut amplifier la variabilité intra-espèces.

Bien que la variabilité saisonnière intra-annuelle de la réponse spectrale puisse contribuer à discerner les espèces ayant une phénologie bien marquée, il peut parfois y avoir des chevauchements entre les cycles phénologiques, en particulier dans les milieux hétérogènes tels que les forêts (Gómez et al., 2016). La variabilité intra-espèces peut parfois être plus importante que la variabilité inter-espèces (Sheeren et al., 2016) et par conséquent, elle ne devrait pas être sous estimée pour une discrimination à l'échelle régionale ou nationale. Ce problème de chevauchement est plus important dans les forêts mixtes⁵ et il est d'autant plus complexe que la résolution spatiale est grossière en raison du mélange de plusieurs espèces au sein d'un même pixel (Ines and Honda, 2005; Bruzzone and Carlin, 2006). Ainsi, des approches multi-temporelles permettant de « généraliser » le comportement spectral intra-espèces sont requises.

D'autres part, pour une cartographie sur de grandes étendues, il faut considérer des échantillons de référence présents en nombre suffisant et représentatifs de la variabilité intra-espèces de chaque classe. Cependant, le principal défi réside dans la collecte de données de référence de qualité, un défi qui croît avec la taille de la région étudiée.

5. Forêts avec un mélange de différentes espèces arborées et/ou avec un sous étage plus ou moins développé

Outre ces difficultés, dès lors qu'il s'agit de grandes étendues, des contraintes purement techniques émergent, en raison des grands volumes de données à gérer et des capacités de traitement et de stockage qu'ils impliquent.

Pour traiter efficacement les images de télédétection sur de grandes étendues et générer des produits cartographiques précis, des stratégies robustes aux changements des conditions d'acquisition et aux variations spatio-temporelles de la réponse spectrale doivent être envisagées.

1.3.2 Approches adoptées pour la cartographie opérationnelle sur de grandes étendues

Apprentissage supervisé pour la classification des images de télédétection

De manière générale, la génération de produits cartographiques à partir d'images de télédétection est réalisée par des approches d'apprentissage automatique supervisées ou non-supervisées. Dans le cas d'un apprentissage supervisé, des données de références sont attribuées aux différentes classes⁶ à identifier. Celles-ci apportent les connaissances de base sur le comportement spectral et temporel de chaque classe, un *a priori* nécessaire pour entraîner un modèle de classification.

Les approches non-supervisées se distinguent des approches supervisées par l'absence de connaissances de base sur les classes. Elles consistent à extraire des groupes (ou *clusters*) présentant des caractéristiques spectrales et/ou temporelles communes, dont l'interprétation est réalisée *a posteriori*. Ces techniques de *clustering* sont principalement utilisées lorsque les références sont limitées ou pour stratifier un échantillonnage de terrain avant de recueillir des références. Malgré le caractère automatique et moins exigeant des approches non supervisées, l'interprétation des *clusters* peut être une tâche difficile, en particulier lorsque la dimension des données est élevée (Chen and Gong, 2013).

La tendance récente des travaux de cartographie sur de grandes étendues a évolué d'une prédominance des techniques non supervisées à une utilisation croissante des approches supervisées (Franklin and Wulder, 2002; Gómez et al., 2016), notamment grâce à l'augmentation de la disponibilité des échantillons de référence à large échelle mais aussi à l'amélioration des performances des algorithmes de classification supervisée (Grekousis et al., 2015).

Outre le choix de l'algorithme de classification, le succès des approches supervisées dépend essentiellement de la qualité des échantillons de référence (Tuia et al., 2016). Ils doivent être précis et représentatifs de la zone à classer. La notion de précision se réfère à la justesse de l'étiquetage ainsi qu'à la datation des données de référence, qui doivent représenter fidèlement la vérité terrain sur la période des acquisitions satellitaires. Par exemple, une peupleraie de 16

6. Classes d'occupation de sol, classes d'essences forestières, etc.

ans au 1^{er} janvier 2017 a de fortes chances d'être coupée au cours de l'année et doit faire l'objet d'une vérification avant d'être utilisée comme référence pour une classification supervisée en 2018. Les données de références doivent également être représentatives de la variabilité spatiale des comportements spectraux au sein de la zone à classifier. La disponibilité d'un nombre suffisant de références actualisées et couvrant toute la diversité est crucial ([Foody and Mathur, 2006](#)). Cependant, cela constitue l'une des principales contraintes des approches de classification supervisée qui est amplifiée dès lors que l'on travaille sur de grandes étendues.

Les références sont habituellement collectées directement sur le terrain, générées par photo-interprétation, ou tirées de bases de données cartographiques globales. Pour une production sur de grandes étendues, les deux premières options sont peu réalistes car il s'agit de tâches fastidieuses et coûteuses, surtout si elles doivent être répétées dans le temps dans le cadre d'approches opérationnelles de production de cartes annuelles. L'ouverture des bases de données publiques et privées de références sur l'occupation des sols et la multiplication des plateformes de production participative de l'information (ou *crowdsourcing*) ont permis de fournir de plus grands volumes de données. Si ces sources d'information sont largement disponibles à faible coût, leurs spécifications sont parfois inadaptées à la tâche d'apprentissage souhaitée, avec par exemple une nomenclature peu riche ou une généralisation cartographique des objets incluant plusieurs classes, mais aussi leur mise à jour et leur qualité ne sont pas toujours assurées, ce qui peut compromettre les performances des modèles de classification supervisée ([McIver and Friedl, 2002](#)). En effet, l'hypothèse qui sous-tend la plupart des approches d'apprentissage supervisé est que les données de référence ont la même distribution des valeurs spectrales, ce qu'on appelle l'espace des caractéristiques (ou *feature space*), que le reste de l'image. Autrement dit, elles sont représentatives de l'image, en incluant toute la variabilité. Dans le cas contraire, la prédiction sur le reste de l'image risque d'être moins performante puisqu'il existe un biais d'échantillonnage (*sampling selection bias*) avec deux espaces de caractéristiques différents (celui des échantillons et celui de la population). Ce décalage entre les données de références utilisées par les algorithmes de classification et la zone à classer, communément connu sous le terme de *datashift*, est à l'origine de la dégradation des performances.

Stratégies pour le passage à l'échelle

Les méthodes de production de données cartographiques à des échelles globales doivent être robustes aux variations spatiales et spectro-temporelles, stables et reproductibles ([Franklin and Wulder, 2002](#)). Inspirés de la littérature sur l'apprentissage automatique, les travaux récents en télédétection ont proposé des approches spécifiques à la classification d'images satellitaires pour la production de cartes thématiques sur de grandes étendues.

Pour la production de cartes thématiques sur de grandes étendues, deux grandes stratégies

sont possibles. La première approche, plutôt intuitive, consiste soit à considérer la zone d'étude comme un ensemble de sites distincts, présentant chacun des conditions locales relativement homogènes, et à créer des modèles de classification locaux adaptés à chaque situation, soit à créer un modèle global prenant en compte toute la variabilité de la zone d'étude. Une telle approche suppose que des références représentatives de la variabilité locale ou globale sont disponibles sur l'ensemble de la zone étudiée. Il s'agit, par exemple, de la stratégie adoptée pour la production de la carte d'occupation des sols à l'échelle de la France du Centre d'Expertise Scientifique de l'Occupation des SOIs (CES OSO) (Inglada et al., 2017), où un modèle de classification est créé pour chaque région éco-climatique pour laquelle des données de référence sont disponibles en quantité suffisante.

La deuxième catégorie d'approches requiert une quantité plus limitée de références en cherchant à généraliser les connaissances dans l'espace ou dans le temps. Le concept de généralisation n'est pas nouveau et il a été historiquement abordé par la communauté à travers des approches d'« extension de signature » ou de « généralisation de signature » (Botkin et al., 1984; Woodcock et al., 2001; Olthof et al., 2005). Ce sont des approches qui consistent à définir des bibliothèques de signatures spectrales pour chaque classe sur une image de référence, puis à stabiliser les définitions à travers d'autres images acquises à des moments différents, dans des sites différents ou même à partir d'autres capteurs. Cela implique soit d'appliquer des méthodes de standardisation de toutes les images, soit de déterminer des classes invariantes dans le temps ou dans l'espace. Cependant, ces approches sont connues pour être sensibles aux variations des conditions d'acquisition (illumination, angle de vue, etc.) et nécessitent malgré tout d'avoir un certain nombre d'échantillons de références couvrant toutes les périodes ou toutes les étendues géographiques (Dannenbergh et al., 2016). Elles ont principalement été mobilisées pour reproduire des classifications sur une base annuelle régulière (McIver and Friedl, 2002; Yin et al., 2020) et peu d'applications montrent leur utilisation sur de grandes étendues (Olthof et al., 2005).

Par ailleurs, l'augmentation de la disponibilité des images satellitaires associée à l'amélioration des diverses résolutions a donné lieu à des espaces de données complexes aux dimensions temporelles, spectrales et spatiales croissantes. L'évolution de la nature et du volume des données de télédétection a permis d'aborder différemment ce problème de généralisation en tirant parti des progrès récents dans les techniques d'apprentissage automatique, notamment des stratégies d'adaptation de domaine (Pan and Yang, 2010; Tuia et al., 2016) qui pourraient s'avérer pertinentes pour la cartographie sur de grandes étendues et que nous allons développer dans la suite.

Apprentissage par transfert : adaptation de domaine

Les stratégies d'adaptation de domaine et, plus généralement d'apprentissage par transfert ont émergé en réponse à ces défis de disponibilité de données de référence de qualité et de *datashift* (Pan and Yang, 2010). L'apprentissage par transfert consiste à transférer des connaissances apprises⁷ sur un premier domaine appelé **source** vers un second domaine appelé **cible**, domaines associés à deux images acquises sur des zones géographiques différentes (transfert spatial) ou sur la même zone mais à de moments différents (transfert temporel).

En fonction de la relation entre les domaines source et cible et la disponibilité des données de références, la littérature définit plusieurs modes de transfert par apprentissage (Pan and Yang, 2010). Dans le cas de la cartographie d'occupation des sols sur de grandes étendues, il s'agit souvent d'un transfert spatial ou temporel d'une même « tâche » correspondant à la procédure de classification supervisée. Dans ce cas, la branche d'apprentissage par transfert qui peut se révéler pertinente et adaptée aux applications de cartographie de l'occupation des sols sur de grandes étendues est celle de l'adaptation de domaine.

Les méthodes d'adaptation de domaine utilisées pour la classification des données de télédétection peuvent être regroupées en quatre catégories (Tuia et al., 2016) :

- **Sélection de caractéristiques invariantes** : cela consiste à réduire l'espace des caractéristiques pour ne conserver que celles qui sont les plus robustes aux changements entre les domaines source et cible. Ces caractéristiques invariantes peuvent être des bandes spectrales ou des attributs qui en sont dérivés. Elles sont décrites comme des points de repère (Gong et al., 2013) car elles conservent leur distribution entre les deux domaines. Le nouvel espace des caractéristiques est ainsi plus stable et l'écart entre les deux domaines est plus réduit.
- **Adaptation de la distribution des données** : il s'agit de créer un nouvel espace de caractéristiques commun aux deux domaines et adapté au classifieur. Des approches d'alignement de distributions ou de calcul de mesures de similarité sont proposées dans la littérature (Petitjean et al., 2011).
- **Adaptation de l'algorithme de classification** : contrairement à la précédente, cette méthode ne s'attaque pas à l'espace des caractéristiques mais plutôt au modèle de classification. Elle consiste à modifier le classifieur entraîné sur le domaine source pour l'adapter au domaine cible en prenant en compte la distribution des données cibles (Bruzzone et al., 2006).
- **Utilisation de l'apprentissage actif** : il s'agit d'un cas particulier de l'approche précédente dans lequel le classificateur, en plus de la distribution, est autorisé à acquérir

7. Connaissances acquises à l'issue d'une tâche d'apprentissage automatique comme par exemple une classification supervisée.

certaines références du domaine cible. Ces références sont sélectionnées en fonction de leur potentiel d'amélioration du modèle de classification initial pour prédire correctement le domaine cible.

Par ailleurs, les publications récentes ont souligné le développement d'approches d'adaptation de domaine profonde (*deep domain adaptation*) en télédétection (S. Garea et al., 2019; Iqbal and Ali, 2020; Tong et al., 2020). Celles-ci combinent la puissance des réseaux neuronaux dans les représentations hiérarchiques des caractéristiques avec les techniques d'adaptation de domaine. Un état de l'art détaillé de ces approches émergentes est fourni dans Wilson and Cook (2020).

Les applications d'adaptation de domaine en télédétection sont en plein essor depuis quelques années explorant différentes approches en utilisant divers classifieurs. Cependant peu de travaux ont exploré les stratégies d'adaptation de domaine dans un contexte opérationnel. Elles ont principalement été appliquées afin de réduire le nombre de données de référence nécessaires à la classification supervisée et plus rarement dans des approches de transfert spatial (Rajan et al., 2008; Bruzzone and Persello, 2009; Alajlan et al., 2014) ou temporel (Bruzzone and Marconcini, 2009; Tardy et al., 2019).

Par exemple, pour la production d'une carte d'occupation des sols avec cinq classes à l'échelle de l'Europe, Alajlan et al. (2014) ont utilisé une approche d'adaptation du classifieur par apprentissage actif à partir d'images MODIS mono-dates. Ils ont montré le potentiel de cette approche pour réduire le nombre de références tout en garantissant de bonnes performances de classification sur une grande étendue géographique.

Dans Tardy et al. (2019), les auteurs ont plutôt appliqué une approche d'adaptation de la distribution des données pour effectuer un transfert temporel entre sept années d'acquisitions multi-temporelles Formosat-2. Ils ont montré une amélioration de la précision globale de la classification de 5 à 20% pour l'identification de 17 classes d'occupation des sols. Toujours dans un contexte de transfert temporel, Bruzzone and Marconcini (2009) ont proposé une approche d'adaptation de l'algorithme de classification pour mettre à jour des cartes d'occupation des sols générées à partir de données Landsat-5 acquises à deux années différentes. Ils ont rapporté une amélioration de 14 à 18% de la précision globale de la classification de cinq classes généralistes d'occupation des sols.

Dans cette section, nous avons montré la pertinence des stratégies d'adaptation de domaine pour le transfert spatial ou temporel des connaissances. Les applications de ces stratégies en télédétection ont très peu abordé le problème de la cartographie sur de grandes étendues d'un point de vue opérationnel. La plupart des travaux sont davantage axés sur des aspects méthodologiques et, par conséquent, les algorithmes sont souvent testés sur de petites zones avec des jeux de données présentant un écart spatial peu marqué, parfois en définissant des

domaines source et cible au sein d'une même emprise d'image, ce qui se traduit par une variabilité plus limitée par rapport à des zones géographiques étendues et très éloignées qui comportent une variabilité supplémentaire, due notamment aux effets biogéographiques. En revanche, les études qui ont été menées à des échelles globales ont généralement privilégié des tâches d'apprentissage simples avec des classes très générales qui sont souvent assez bien discriminées (eau, urbain, végétation, etc.). Par ailleurs, dans ces approches, les dimensions spectrale et temporelle n'ont jamais été prises en compte conjointement dans un contexte opérationnel de production cartographique sur de grandes étendues.

1.4 Positionnement de la thèse

Dans ce premier chapitre, nous avons vu qu'il n'existe pas de méthodes spécifiques à la reconnaissance automatique des plantations de peuplier par télédétection. Les travaux antérieurs se sont plutôt focalisés sur les peupliers naturels dans des massifs forestiers. Les approches méthodologiques développées dans le cadre des travaux d'identification des essences forestières ont généralement été menées localement et leur généralisation sur de grandes étendues soulève des problèmes méthodologiques qu'il convient de traiter.

1.4.1 Problématique et questions de recherche

Ainsi, la problématique de cette thèse est de développer une démarche méthodologique robuste permettant de reconnaître les plantations de peupliers sur de grandes étendues tout en tenant compte des contraintes impliquées par la large échelle, principalement liées aux acquisitions asynchrones et à la non-stationnarité spatiale de la réponse spectrale. Le contexte actuel de disponibilité des données Sentinel-2 avec une couverture globale et de hautes résolutions spatiale, spectrale et temporelle, et de développement d'algorithmes de classification et de généralisation des connaissances, apparaît favorable pour réinterroger les capacités de la télédétection à appréhender cette problématique. Les principales questions auxquelles cette thèse se propose de répondre sont les suivantes : Quel est l'apport de l'information temporelle fournie avec les acquisitions Sentinel-2 pour l'identification des peupleraies ? Existe-t-il une fenêtre spectrale permettant de mieux discriminer les peupleraies des autres essences ? Comment faire face au nombre d'échantillons limités pour une classification supervisée à large échelle ? Comment tenir compte du bruit et de la variabilité spatiale et temporelle des échantillons de référence disponibles ? Enfin, comment prendre en compte la diversité de configurations des peupleraies dans un contexte opérationnel de cartographie à l'échelle nationale ?

1.4.2 Objectifs et démarche suivie

Le principal objectif de la thèse consiste à étudier le potentiel des séries temporelles Sentinel-2 pour distinguer les peupleraies des autres essences en tenant compte des différents contextes populicoles avec notamment des pratiques variées et différents stades de développement. Il s'agit alors de relever l'un des enjeux majeurs en télédétection forestière, à savoir la large échelle avec toutes les contraintes inhérentes. Cette démarche implique l'utilisation d'approches robustes inspirées des nouvelles avancées en apprentissage automatique avec les techniques d'adaptation de domaine. Au-delà de l'aspect méthodologique, une contrainte importante de la contribution attendue est le caractère opérationnel de la méthode pour la filière populicole, ce qui implique des contraintes en termes de facilité de mise en œuvre en vue de permettre une production annuelle systématique.

Pour répondre aux différents enjeux présentés, la démarche proposée est la suivante :

1. D'abord, nous avons étudié la capacité des données Sentinel-2 à discriminer les peupleraies des autres essences à des échelles locales représentées par trois sites distincts offrant chacun des conditions relativement homogènes. À ce stade nous avons exploré toute l'information spectrale et temporelle disponible sans chercher à identifier celle qui est la plus pertinente pour les peupleraies.
2. Pour le passage à l'échelle, nous avons cherché à tester le transfert de connaissances depuis un site source vers un site cible. Nous avons exploré une première méthode d'adaptation de domaine par adaptation des distributions, le transport optimal, très peu utilisée en télédétection. Nous avons testé une première variante non supervisée ne nécessitant pas d'étiquettes sur le site cible, puis nous l'avons adoptée en mode semi-supervisé.
3. Une autre approche d'adaptation de domaine a ensuite été envisagée au regard des limites de la première. Il s'agit d'une approche semi-supervisée d'adaptation du classifieur par apprentissage actif. Elle présente l'avantage d'être performante avec peu de références pour apprendre.
4. Une alternative a été proposée à l'approche par apprentissage actif pour des contraintes d'opérationnalité. Il s'agit d'une démarche d'apprentissage plus classique mais qui s'appuie sur un nouvel indice spectral proposé, défini spécifiquement pour reconnaître automatiquement les peupleraies.
5. La dernière approche a été retenue et adoptée pour la production cartographique à l'échelle de la France métropolitaine aboutissant à la première carte nationale des peupleraies dérivée d'images satellitaires.

Le point 1 ne fait pas l'objet d'un chapitre spécifique dans le manuscrit mais sera abordé

plus en détail lors du développement du point 3. En revanche, les autres points (2 à 5) seront traités dans des chapitres indépendants précisés dans la section suivante.

1.4.3 Organisation du manuscrit

Le manuscrit de cette thèse s’articule en cinq parties :

- La première partie (I) introductive comporte ce premier Chapitre 1 qui présente le contexte général de la thèse et parcourt l’état de l’art sur les thématiques abordées, notamment les méthodes existantes pour la cartographie des essences forestières à large échelle par télédétection.
- La deuxième partie (II) comprend deux chapitres. Le premier (Chapitre 2) est consacré à la description des sites d’études, des données utilisées ainsi que les pré-traitements appliqués qui sont communs à toutes les approches développées dans la thèse. Le Chapitre 3 détaille la méthodologie générale qui sera utilisée tout au long de la thèse. Les choix des algorithmes d’apprentissage automatique, ainsi que les méthodes d’optimisation et d’évaluation seront présentés.
- La troisième partie (III) porte sur l’utilisation de l’adaptation de domaine pour la cartographie des peupleraies à large échelle. Le Chapitre 4 aborde une première technique d’adaptation des distributions des données. Le Chapitre 5, présenté sous forme d’article, propose une seconde approche d’adaptation du classifieur par apprentissage actif.
- La partie IV comprend également deux chapitres. Le Chapitre 6 qui constitue la troisième approche proposée dans cette thèse pour le passage à l’échelle. Il est présenté sous forme d’article. Le Chapitre 7 décrit les moyens déployés pour la production cartographique à l’échelle nationale.
- Enfin la dernière partie (V) qui est divisée en deux chapitres (8 et 9), résumera les contributions de la thèse et discutera des principaux enjeux et perspectives.

La structure du manuscrit est résumée dans le schéma de la Figure 1.13.

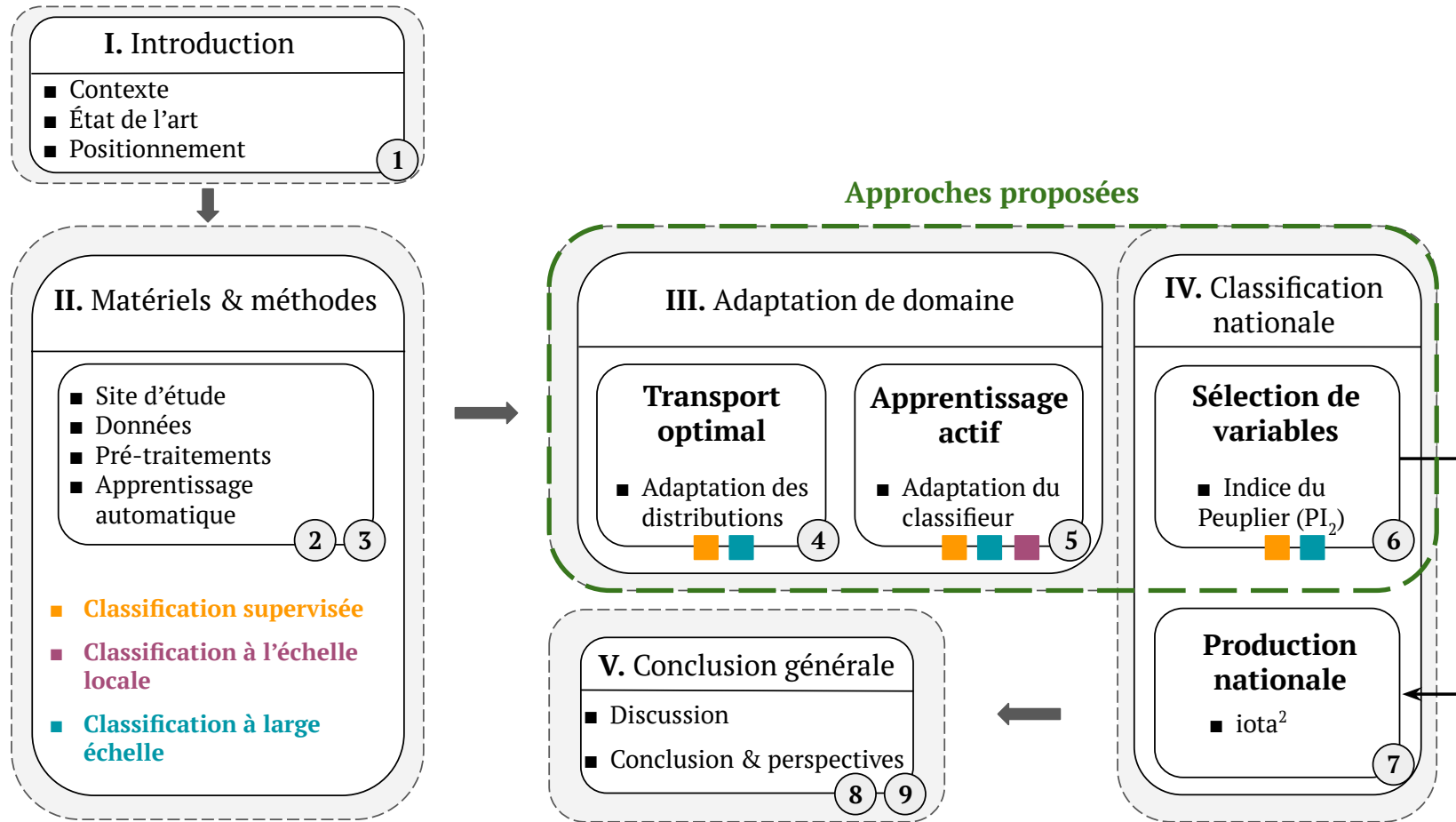


FIGURE 1.13 – Schéma de la structure de la thèse (adapté depuis Almeida (2017)). Les cercles gris indiquent les numéros des chapitres concernés.

Deuxième partie

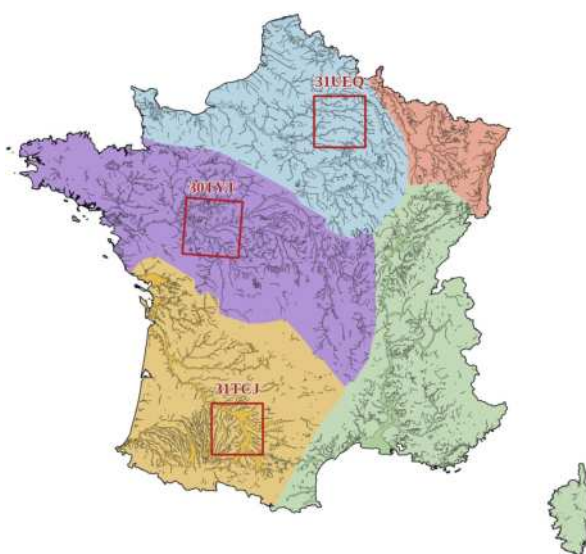
Matériels et méthodes

Chapitre 2

Sites d'étude, données et pré-traitements

Sommaire

| | | |
|-------|---|-----------|
| 2.1 | Choix des sites populicoles | 41 |
| 2.2 | Données satellitaires | 43 |
| 2.2.1 | Caractéristiques des données Sentinel-2 | 43 |
| 2.2.2 | Pré-traitements des données Sentinel-2 | 44 |
| 2.2.3 | Construction de séries temporelles annuelles homogènes : résolution spatiale, données manquantes et asynchrones | 47 |
| 2.3 | Données de référence | 50 |
| 2.4 | Stratégie d'échantillonnage | 52 |
| 2.4.1 | Stratification par classe | 52 |
| 2.4.2 | Détection des valeurs aberrantes | 52 |



La première partie de ce chapitre est consacrée à la présentation des données utilisées dans ces travaux de thèse, à savoir les images satellitaires et les données de référence. La deuxième partie décrira les pré-traitements appliqués afin d'extraire les données d'entrée pour la démarche d'apprentissage automatique proposée.

2.1 Choix des sites populicoles

Afin de couvrir la variabilité des peupleraies françaises, trois sites populicoles contrastés en termes de conditions climatiques, de modes de gestion sylvicoles, et de types de cultivars ont été sélectionnés. Ils sont localisés dans la région Grand-Est, la région Centre-Val de Loire et dans le Sud-Ouest. Ils sont inclus dans deux Grandes Régions ÉCOlogiques (GRECO) et trois SylvoécORégions (SER) d'alluvions différentes. En effet, en fonction des facteurs biogéographiques déterminant la production forestière et la répartition des grands types d'habitats forestiers, l'IFN a découpé le territoire métropolitain en 86 SER. Ces SER sont regroupées en 11 GRECO où chacune présente des conditions homogènes et discontinues avec les GRECO adjacentes.

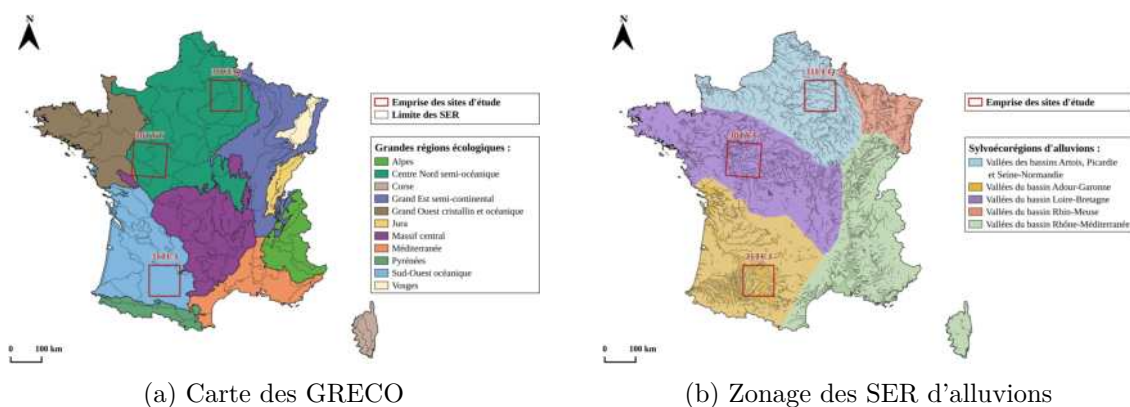


FIGURE 2.1 – Localisation des sites d'étude sur la carte des GRECO et en fonction du zonage des SER d'alluvions.

Les trois sites de référence (représentés par des carrés rouges dans la Figure 2.1) se trouvent sur sept SER regroupées au sein de deux GRECO :

- **Grand Ouest cristallin et océanique** : elle est caractérisée par un climat océanique et frais avec une forte hygrométrie. Les pluies sont abondantes et régulières et les températures sont plutôt douces. Les forêts représentent 11% et elles sont réparties de manière inégale sur l'ensemble de la GRECO. (IGN, 2013).
- **Sud-Ouest océanique** : cette grande région écologique est caractérisée par un climat océanique généralement doux, mais qui se dégrade vers l'est avec l'apparition d'une sécheresse estivale. Les précipitations sont saisonnières mais plus concentrées en automne

et en hiver. Le taux de boisement est de 30% avec des vallées fluviales du centre au sud de la région où sont plantées de nombreuses peupleraies (IGN, 2013).

Un deuxième découpage complémentaire au premier a permis de diviser le territoire en fonction des grands bassins hydrologiques correspondant à cinq SER d'alluvions aux enjeux populicoles et patrimoniaux importants (IFN, 2011). L'emprise des trois zones populicoles de référence correspond à celles des acquisitions Sentinel-2, fournies sous forme de tuiles carrées de 100 km × 100 km. La localisation des trois tuiles sur la carte des GRECO et selon le zonage¹ des SER d'alluvions, tels que définis par l'IFN, est présentée dans la Figure 2.1.

Par ailleurs, les trois sites de référence sont également représentatifs de la diversité des pratiques populicoles, qui sont principalement liées à la disponibilité en eau. En effet, dans la région du Sud-Ouest, la pluviométrie relativement faible combinée à une sécheresse estivale de plus en plus fréquente rendent le travail du sol indispensable dans les peupleraies afin de limiter la compétition pour l'eau tout au long de leur cycle de développement. Par conséquent, le déploiement d'un sous-étage est très limité (Figure 2.3a). À l'inverse, dans la région Est, les précipitations sont plus abondantes et le climat est plus doux. Un sous-étage herbacé, voire arbustif, est souvent présent dans la région et le travail de sol reste très rare (Figure 2.3b). Avec des conditions intermédiaires entre les deux régions précédentes, dans le bassin populicole du Centre, le travail du sol est généralement pratiqué pendant les cinq ou six premières années de croissance des peupleraies (durée variable selon les cultivars plantés), mais une fois l'arbre bien développé, le sol est beaucoup moins entretenu et un sous-étage est généralement présent jusqu'à la fin du cycle (Figure 2.3c).

Le choix de ces trois sites de référence est également stratégique par rapport à la localisation des trois grands bassins populicoles où se concentrent les communes ayant les plus importantes surfaces de plantation (Figure 2.2).

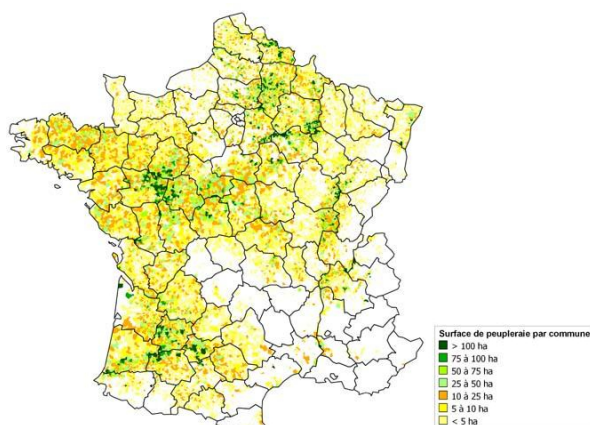


FIGURE 2.2 – Carte des surfaces de peupleraies par commune dérivées du cadastre 2003. Source : <https://www.emballage-leger-bois.fr/node/12>

1. Les SER d'alluvions sont censées être azonales, mais ont elles été délimitées pour plus de lisibilité.



(a) Peupleraie dans le Sud-Ouest



(b) Peupleraie dans le Grand-Est



(c) Peupleraie dans le Centre

FIGURE 2.3 – Illustrations de peupleraies typiques dans les trois sites de référence. En (a) une peupleraie du Sud-Ouest sans sous-étage, en (b) une peupleraie caractéristique du Grand-Est avec une strate ligneuse en sous-étage et en (c) une peupleraie mature dans le Centre avec une strate herbacée dense de mégaphorbiaies. Photos ©Eric Paillassa.

2.2 Données satellitaires

2.2.1 Caractéristiques des données Sentinel-2

Les données satellitaires utilisées dans ces travaux de thèse sont issues de la mission Sentinel-2 lancée dans le cadre du programme Copernicus de l'ESA. Celle-ci consiste en une constellation de deux satellites jumeaux Sentinel-2A et Sentinel-2B mis en orbite respectivement en juin 2015 et mars 2017. Phasés à 180° , ils permettent à deux d'observer la surface de la Terre avec une répétitivité de cinq jours (10 jours avec un seul satellite).

Le capteur *MultiSpectral Instrument* (MSI) embarqué à bord permet l'acquisition d'images dans 13 bandes spectrales allant du visible au moyen infrarouge à une résolution spatiale de 10 à 60 m sur une fauchée large de 290 km. Dans le cadre de ces travaux, seules les 10 bandes spectrales de 10 et 20 m de résolution spatiale ont été utilisées. Les bandes 1 (443 nm), 9 (945 nm) et 10 (1375 nm) à 60 m sont dédiées aux pré-traitements, notamment la correction atmosphérique et le filtrage des nuages. Les résolutions spatiale et spectrale des 13 bandes sont illustrées dans la Figure 2.4.

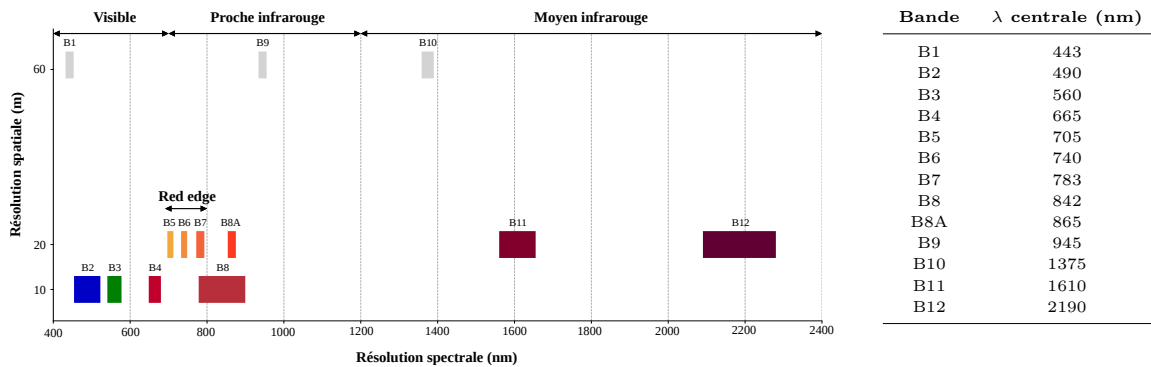


FIGURE 2.4 – Résolutions spatiale (taille du pixel au sol) et spectrale (largeur de bande) des images Sentinel-2.

L'ESA distribue les images Sentinel-2 sous forme de tuiles de $100 \text{ km} \times 100 \text{ km}$ suivant le système de carroyage militaire *Military Grid Reference System* (MGRS) dérivé du système de coordonnées UTM. Les images sont fournies au niveau 1C qui est le niveau standard de distribution des données satellitaires. Il s'agit de produits de réflectance au sommet de l'atmosphère ou *Top Of Atmosphere* (TOA) où les acquisitions sont orthorectifiées avec un Modèle Numérique de Terrain (MNT) puis découpées en tuiles de 100 km^2 (ESA, 2015). La France est couverte par 90 tuiles Sentinel-2 qui sont acquises lors du passage du satellite autour d'une ou plusieurs orbites relatives (Figure 2.5).

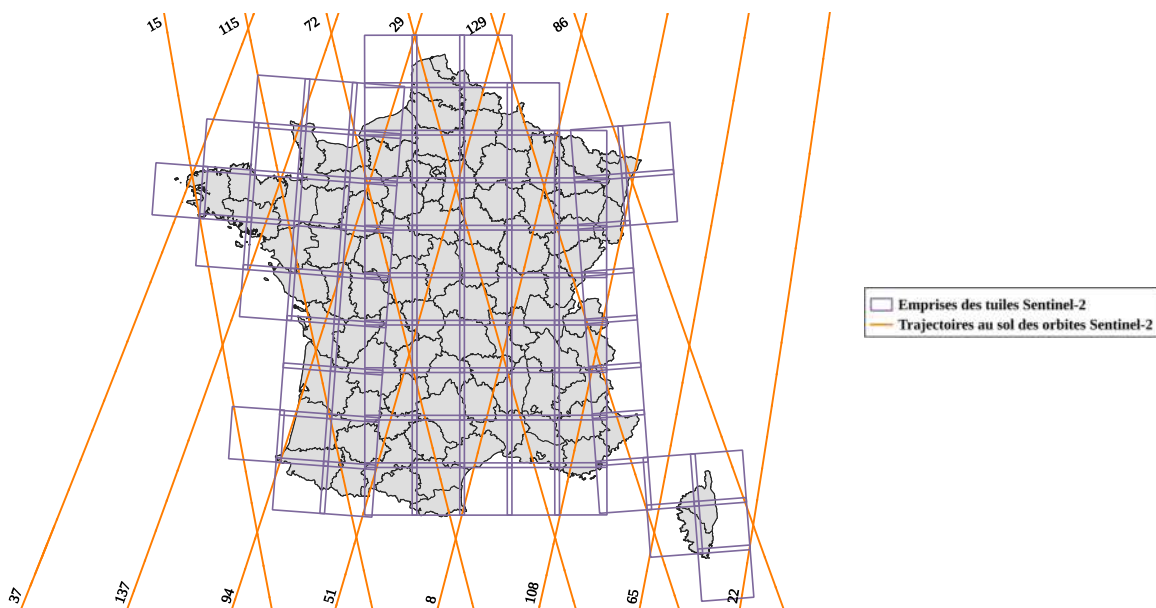


FIGURE 2.5 – Illustration des tuiles Sentinel-2 et des trajectoires au sol des orbites relatives passant par la France.

2.2.2 Pré-traitements des données Sentinel-2

Au moment de l'acquisition, le signal enregistré par Sentinel-2 n'est pas homogène sur toutes les surfaces observées en raison des perturbations atmosphériques (présence d'aérosols,

molécules d'air, nuages, etc.) mais également de la géométrie d'acquisition (angle de vue, relief, etc.). Une série de corrections est donc appliquée afin d'obtenir des valeurs de réflectance comparables entre les différentes acquisitions. Ces opérations sont pour la plupart effectuées par les plateformes de distribution des données satellitaires. Dans le cas des données Sentinel-2 utilisées dans ces travaux de thèse, ces corrections sont appliquées par le Centre National d'Études Spatiales (CNES) en s'appuyant sur la chaîne de traitement MAJA² (Hagolle et al., 2017). Les données Sentinel-2 pré-traitées sont distribuées par la plateforme du pôle des données et de services surfaces continentales Theia. L'ensemble de ces pré-traitements visant à convertir les produits de niveau 1C en produits de niveau 2A seront décrits en détail dans la suite.

Détection des nuages et des ombres

Dans la chaîne de traitement MAJA, la première étape consiste à détecter les pixels nuageux. Cette détection comprend tout d'abord l'identification des nuages hauts grâce à la bande *cirrus* de Sentinel-2 (bande 10 à 1375 nm) qui correspond à une forte absorption atmosphérique. Ensuite, comme les nuages denses et opaques sont caractérisés par une forte réflectance dans le bleu (bande 2 à 490 nm), la deuxième étape repose sur un critère multi-temporel qui repère une augmentation soudaine de la réflectance dans le bleu d'une image à l'autre, signe de la présence d'un nuage. Afin d'éviter les sur-détections, chaque pixel potentiellement nuageux dans une image est comparé aux images précédentes. Comme il y a une faible probabilité que plusieurs nuages successifs soient similaires et situés au même endroit, tous les pixels ayant un coefficient de corrélation supérieur à 0.9 avec l'une des images sont écartés (Baetens et al., 2019). Une fois les nuages détectés, la position de leur ombre ainsi que leur altitude sont estimées et projetées sur l'image pour générer un masque des ombres.

Correction radiométrique

Cette étape consiste à transformer le signal brut enregistré par le satellite en une grandeur physique homogène, la réflectance *Top Of Canopy* (TOC), qui reflète les valeurs les plus proches de la surface du sol après élimination des effets de la couche atmosphérique.

Avant d'être enregistrée par le satellite, la lumière solaire est perturbée par sa traversée atmosphérique qui est fonction de plusieurs facteurs, notamment l'épaisseur optique des aérosols et la quantité de vapeur d'eau. Ces paramètres ne sont pas constants sur toutes les surfaces éclairées. En outre, en fonction de sa position, le soleil n'éclaire pas les cibles sous le même angle. En plus du champ de vue du capteur optique, le signal enregistré au niveau du

2. Logiciel développé en coordination entre le CNES/Centre d'Études Spatiales de la BIOSphère (CESBIO) et le *German Aerospace Center* (DLR) utilisant la méthode *Multi-sensor Atmospheric Correction and Cloud Screening* (MACCS).

satellite est influencé par l'ensemble de ces perturbations. Il correspond au flux réfléchi par unité de surface du capteur et il est exprimé en valeurs brutes, appelées comptes numériques.

La première étape de la correction radiométrique consiste tout d'abord à normaliser le flux réfléchi par rapport à la surface du détecteur. Cette normalisation est généralement effectuée à partir de relations d'étalonnage permettant de convertir les comptes numériques en une grandeur physique indépendante du détecteur, la luminance. Afin de s'abstraire de l'éclairement solaire, qui est fonction de la distance entre le Soleil et la Terre, de l'angle zénithal et de la quantité d'énergie relative provenant du Soleil, la luminance est ensuite normalisée en une valeur de réflectance TOA. Enfin, la dernière étape consiste à éliminer les effets atmosphériques en passant d'une réflectance TOA en une réflectance TOC à partir d'informations précises sur les conditions atmosphériques telles que l'épaisseur optique des aérosols, la quantité de vapeur d'eau, etc.

Dans la chaîne de traitement MAJA, un ensemble de critères multi-temporels et multi-spectraux sont combinés afin d'estimer l'épaisseur optique des aérosols (Hagolle et al., 2015). De même, grâce à la présence de la bande 9 (945 nm) située dans une forte bande d'absorption de la vapeur d'eau, la quantité de la vapeur d'eau est estimée. Une fois ces informations déterminées, les effets atmosphériques sont corrigés et les réflectances TOC sont calculées.

Correction géométrique

Les images acquises par les satellites subissent un certain nombre de distorsions géométriques principalement dues à des erreurs de positionnement du satellite sur son orbite (mouvement de la plateforme et sa vitesse, etc.), la courbure et la rotation de la Terre pendant l'acquisition ou des effets du relief. Avant correction, les images ne sont donc pas superposables et ne possèdent pas de coordonnées géographiques. Les corrections géométriques permettent de réduire l'ensemble de ces déformations et d'avoir une représentation la plus fidèle du terrain. Cela consiste à appliquer des modèles géométriques qui prennent en compte la position du satellite, son altitude, etc. afin de rectifier géométriquement les images et leur attribuer des coordonnées au sol (orthorectification). Les données distribuées par Theia sont orthorectifiées et également corrigées des effets de pente, qui impliquent des variations d'éclairement dues au relief, en s'appuyant sur le MNT SRTM à 90 m de résolution spatiale (Hagolle et al., 2017).

Les acquisitions Sentinel-2 distribuées par Theia sont donc au niveau de traitement 2A : des images en réflectance TOC, corrigées des effets atmosphériques, géométriques et des effets de pente (produits *Flat REflectance* (FRE))³ et fournies avec un masque des nuages et de

3. Theia fournit les produits *Surface REflectance* (SRE) corrigées des effets atmosphériques mais sans correction des effets de pente et les produits FRE qui ont subi une correction supplémentaire des effets de pente.

leurs ombres ainsi qu'un masque des valeurs nulles et des pixels saturés (Hagolle et al., 2015). Les images Sentinel-2 pré-traitées par Theia sont encodées en 16 bits au format entier où les valeurs de réflectance de surface sont multipliées par 10 000.

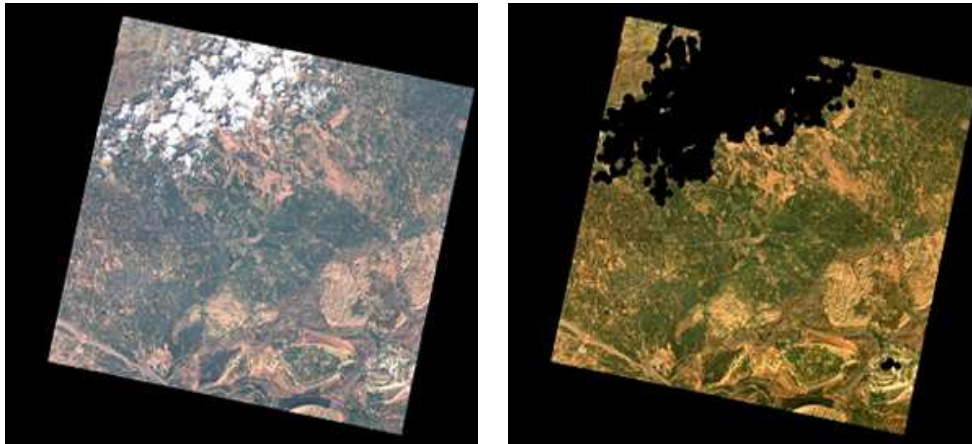


FIGURE 2.6 – Illustration des deux niveaux de traitements 1C et 2A. En (a) une image du 16/03/2006 en entrée de la chaîne de traitement MAJA et en (b) la même image au niveau 2A après correction atmosphérique et géométrique et détection des nuages (affichés en noir). Source : <https://labo.obs-mip.fr/wp-content-labo/uploads/sites/19/2018/08/Produits.png>

Dans le cadre de cette thèse, les données Sentinel-2 pré-traitées sur les trois tuiles de référence ont été téléchargées pour les deux années 2017 et 2018. Selon la partie abordée, tout ou partie de ces données sont exploitées.

2.2.3 Construction de séries temporelles annuelles homogènes : résolution spatiale, données manquantes et asynchrones

Ré-échantillonnage spatial

Pour travailler sur des images à différentes dates et avec différentes résolutions spectrales, il est nécessaire d'avoir des tailles de pixels équivalentes avec un même niveau d'information en entrée des procédures d'apprentissage automatique.

Étant donné que les 10 bandes spectrales Sentinel-2 n'ont pas la même résolution spatiale (équivalente ici à la taille des pixels), il est indispensable de procéder à un ré-échantillonnage spatial des pixels de 20 m à 10 m (sur-échantillonnage) ou de 10 m à 20 m (sous-échantillonnage). Afin de préserver le maximum d'informations spatiales disponibles, les six bandes de 20 m ont été ré-échantillonnées à 10 m. La méthode la plus courante a été utilisée afin de diviser les pixels de 20 m en quatre pixels de 10 m en leur attribuant la même valeur que le pixel original.

Comblement des lacunes

Comme décrit dans la partie 1.2.1, les longueurs d'onde dans lesquelles opèrent les capteurs optiques tels que Sentinel-2 sont influencées par l'épaisseur atmosphérique. En effet, la présence d'une couverture nuageuse rend l'atmosphère plus ou moins opaque, les nuages diffusent le rayonnement réfléchi par les cibles ce qui ne permet pas d'avoir un signal exploitable au niveau du capteur. Les corrections atmosphériques ne peuvent en aucun cas corriger leur présence puisque le signal ne parvient pas à les traverser.

Par conséquent, d'une date à l'autre et d'un endroit à l'autre, en fonction des conditions météorologiques, les acquisitions peuvent être incomplètes (pixels invalides dans l'image, sans valeur de réflectance). Un exemple du nombre de fois où un pixel Sentinel-2 n'est pas masqué en 2018 (pixel valide qui ne correspond pas à un nuage ou à une ombre ni à une valeur saturée) est présenté dans la Figure 2.7 sur l'ensemble de la France. Théoriquement, avec une résolution temporelle de cinq jours, chaque endroit est observé environ 73 fois par Sentinel-2. Mais en réalité, tous les pixels ne sont pas vus avec le même nombre de fois. En raison du recouvrement entre les trajectoires orbitales, certaines régions sont imagées tous les trois jours, ce qui donne plus de 100 observations par an (bandes vertes et jaunes). À l'inverse, les conditions climatiques défavorables dans certaines régions font que le nombre d'acquisitions est plus faible (bandes rouges et orange).

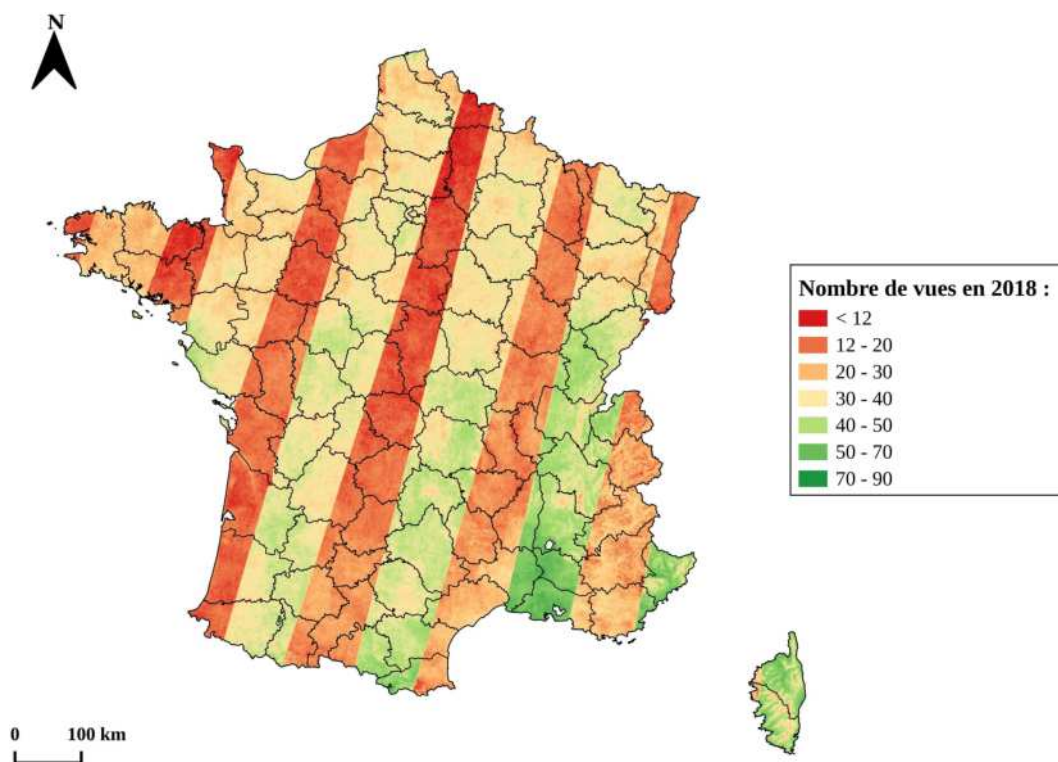


FIGURE 2.7 – Nombre de fois où les pixels Sentinel-2 sont vus (ou nombre de vues) en France pour l'année 2018. La succession des différentes bandes de couleur est liée à la trajectoire au sol des orbites Sentinel-2 (Figure 2.5).

L'une des façons de traiter ce problème de données incomplètes est de procéder à un comblement des lacunes ou *gap-filling*. Les approches visant à combler les lacunes utilisent des modèles mathématiques dont les paramètres sont déterminés à partir des observations valides disponibles. Ces modèles sont ensuite appliqués pour estimer les valeurs des observations invalides (Liu et al., 2017a). Il existe une multitude d'approches qui s'appliquent de manière locale ou globale, notamment celles qui définissent un filtre pour combler les lacunes avec des observations valides dans une fenêtre glissante ou celles qui reconstruisent les données par interpolation. Plusieurs travaux ont évalué l'efficacité de ces approches dans différents contextes (Julien and Sobrino, 2010; Atkinson et al., 2012; Kandasamy et al., 2013).

Parmi les différentes méthodes proposées dans la littérature, c'est l'interpolation linéaire multi-temporelle qui a été retenue pour sa simplicité et sa capacité à conserver autant que possible les valeurs originales des pixels. Cela consiste à reconstruire les séries temporelles en interpolant les valeurs manquantes (pixels masqués) à partir des données valides de la série. En effet, pour une date donnée, si un pixel n'est pas détecté comme nuageux dans le masque de nuages, sa valeur de réflectance est conservée sinon elle est interpolée à partir des dates sans nuages antérieures et postérieures les plus proches (Inglada et al., 2015). Elle permet ainsi de générer des séries temporelles sans lacunes en introduisant un minimum de nouvelles informations. L'efficacité de l'interpolation linéaire a déjà été démontrée, en particulier avec des séries temporelles à forte répétitivité comme Sentinel-2 (Julien and Sobrino, 2010; Kandasamy et al., 2013). Elle offre un bon compromis entre le temps de calcul et la précision des algorithmes de classification (Kandasamy et al., 2013; Inglada, 2016). Une exemple d'un profil temporel reconstitué pour un pixel de peuplier est fourni en Annexe A.2.

Cependant, l'efficacité de cette approche est fortement liée à la qualité du masque des nuages et des ombres. En effet, si un pixel invalide n'est pas détecté dans le masque, sa valeur n'est jamais interpolée. Cela peut avoir un double impact sur le comblement des lacunes dans le reste de la série temporelle car si le pixel est considéré comme valide, sa valeur peut être utilisée pour interpoler d'autres pixels détectés comme invalides. Certains travaux ont montré que cela peut avoir un impact sur les performances de la classification principalement si des classificateurs peu robustes au bruit sont impliqués (Karasiak et al., 2019).

D'autres méthodes d'interpolation telles que les splines cubiques ont également été proposées avec les données Sentinel-2 mais elles sont plus coûteuses en temps de calcul et semblent moins efficaces en termes de performance de classification que l'interpolation linéaire (Inglada et al., 2015).

Ré-échantillonnage temporel

Au delà des contraintes d'acquisitions, Sentinel-2, comme la plupart des satellites optiques, est placé sur une orbite héliosynchrone lui permettant d'observer systématiquement la même surface toujours à la même heure solaire locale en gardant un angle constant. Cependant, toutes les surfaces de la Terre ne sont pas observées en même temps, et les acquisitions suivent un cycle orbital⁴ de 143 orbites relatives pour un cycle complet Sentinel-2⁵, dont environ 14 qui traversent la France (Figure 2.5). Par conséquent, sur de grandes étendues, les acquisitions sont asynchrones et les séries temporelles d'images peuvent être irrégulières d'une région à l'autre, même après le comblement des données manquantes.

Lorsque l'on travaille sur une zone d'étude spécifique impliquant une seule tuile, cela ne pose pas de problème. En revanche, dès lors qu'on cherche à construire un modèle de classification global qui permet la classification de différentes tuiles sur une large étendue, c'est un obstacle à surmonter car il faut s'assurer d'avoir le même nombre de variables spectrales quel que soit l'endroit. Ainsi, pour s'affranchir de ce problème de dimensions, nous avons utilisé une stratégie consistant à ré-échantillonner toutes les séries temporelles selon une grille temporelle commune par interpolation linéaire. C'est une approche simple qui permet d'homogénéiser temporellement toutes les données sans affecter les performances de classification. Elle s'inspire de celle adoptée pour construire la carte nationale d'occupation des sols du CES OSO (Inglada et al., 2015).

Le ré-échantillonnage des dates avec une grille temporelle commune a été réalisé en même temps que l'interpolation linéaire pour combler les lacunes en utilisant le module *Temporal gapfilling for image time series* de la bibliothèque Orfeo Toolbox (Inglada, 2016). Il a été effectué sur les séries temporelles d'images Sentinel-2 de 2017 et 2018 des trois tuiles d'études. Les dates des acquisitions et les dates interpolées sont présentées dans les Tableaux A.1 et B.1 en annexe.

2.3 Données de référence

Les algorithmes d'apprentissage automatique nécessitent de disposer de données de référence sur lesquelles repose le mécanisme d'apprentissage automatique. Celles-ci sont soit dérivées de la BD Forêt V2 décrite dans la section 1.1.3, soit créées par photo-interprétation.

Parmi les 32 types de formations végétales recensés par la BD Forêt V2, seules les classes de feuillus ont été considérées pour les trois tuiles d'études. Elles sont renseignées dans le Tableau 2.1.

4. Il est confondu avec le cycle de revisite pour les satellites héliosynchrones à angle d'observation constant.

5. https://www.esa.int/ESA_Multimedia/Videos/2016/08/Sentinel-2_global_coverage

TABLEAU 2.1 – Classes de feuillus extraites sur les trois sites d'étude.

| Type de Formation Végétale (TFV) | Code TFV | Essence |
|--|---------------------------------|----------------|
| Peupleraie | FP | Peuplier |
| Forêt fermée de robinier pur | FF1-14-14 | Robinier |
| Forêt fermée de châtaignier pur | FF1-10-10 | Châtaignier |
| Forêt fermée de chênes décidus purs | FF1G01-01 | Chênes décidus |
| Forêt fermée de hêtre pur | FF1-09-09 | Hêtre |
| Forêt fermée de feuillus purs ou en îlots, autre feuillu pur, ou en mélange | FF1-00, FF1-49-49, FF1-00-00 | Feuillus |
| Forêt ouverte de feuillus purs | FO1 | Feuillus |

En raison du décalage entre les millésimes de la BD Forêt (2004 à 2015) et les années de prise de vues Sentinel-2 (2017 et 2018), les références de feuillus pourraient potentiellement ne plus être à jour. Avec un cycle de rotation de 15 ans, les peupleraies sont les plus susceptibles d'avoir subi un changement. Par conséquent, deux jeux d'échantillons ont été construits. Un premier jeu formé par des polygones des sept classes sélectionnés depuis la BD Forêt et un deuxième jeu où seuls les polygones de peuplier ont été mis à jour par photo-interprétation pour les deux années d'étude 2017 et 2018. L'objectif étant de quantifier l'impact de la mise à jour des données sur la performance des modèles. Cette vérification a été effectuée à partir d'orthophotographies IRC (si disponibles), d'images Sentinel-2 ainsi que d'images Google Earth et elle a permis par ailleurs de créer de nouvelles références de peuplier.

Les polygones de peuplier photo-interprétés correspondent à des plantations de trois ans ou plus. Lorsque les arbres sont plus jeunes, le couvert végétal peut être partiel et les peupleraies risquent d'être confondues avec d'autres plantations (Figure 2.8).

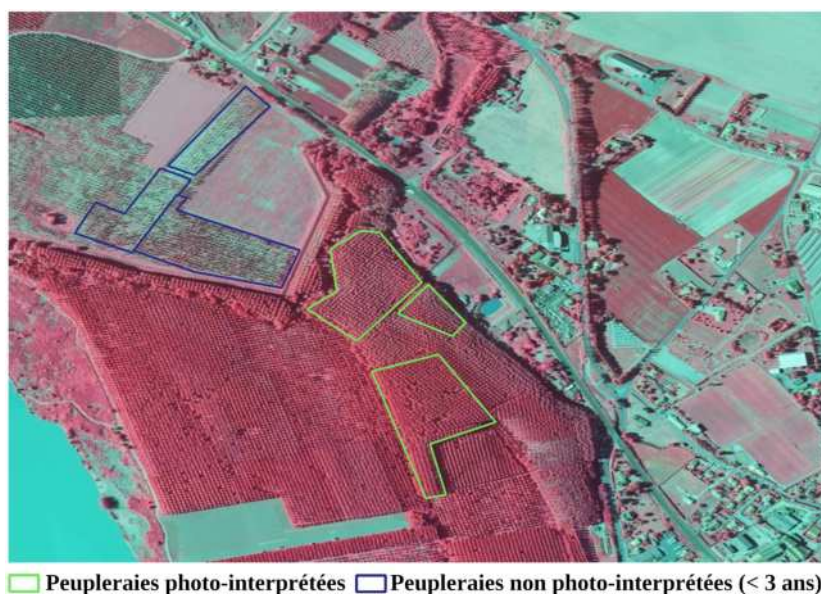


FIGURE 2.8 – Exemple de sélection de peupleraies pour la photo-interprétation sur un fond d'orthophotographies IRC de 2017.

Au total 230 et 300 polygones de peuplier ont été photo-interprétés pour les années 2017 et 2018. Une validation sur le terrain avec les partenaires forestiers du CNPF a confirmé la photo-interprétation des peupleraies de 2017, où seulement deux polygones correspondaient à des plantations d'eucalyptus, ce qui représente moins de 1% d'erreur d'identification.

2.4 Stratégie d'échantillonnage

Après avoir corrigé et harmonisé les données satellitaires, il convient de constituer l'espace des caractéristiques en entrée des algorithmes d'apprentissage automatique. Cela consiste à extraire pour chaque polygone de référence les informations correspondantes dans chacune des séries temporelles d'images (par tuile et par année) ainsi que la classe d'essence qui lui est associée (une des sept classes considérées). L'échelle de travail étant le pixel, ces valeurs seront extraites pour tous les pixels se retrouvant à l'intérieur des polygones de référence. Ces pixels constituent donc les échantillons de référence.

2.4.1 Stratification par classe

Tous les échantillons de référence ne seront pas utilisés en entrée des procédures d'apprentissage. Ils sont divisés en deux sous-ensembles égaux : 50% pour les échantillons d'entraînement qui serviront à l'apprentissage automatique et 50% pour les échantillons de test qui seront utilisés à l'issue de cette étape pour une évaluation indépendante des performances.

Afin que l'évaluation soit la plus fiable et la plus indépendante possible, et pour éviter l'influence des pixels voisins appelée auto-correction spatiale (Karasiak et al., 2019), la séparation des deux sous ensemble d'échantillons se fait sur la base des polygones de référence. En effet, tous les pixels d'entraînement appartiennent à des polygones différents de ceux des pixels de test.

2.4.2 Détection des valeurs aberrantes

Certains échantillons de référence peuvent contenir des valeurs aberrantes à certaines dates qui peuvent provenir soit des séries temporelles soit des données de référence elles-mêmes.

Dans les séries temporelles reconstruites et harmonisées, il existe parfois des nuages résiduels non détectés ou encore de voiles nuageux non identifiés dans le masque. Dans ce cas, les pixels concernés ne sont pas corrigés lors de l'étape du *gap-filling* (car ils sont considérés comme valides puisqu'ils sont absents du masque) et conservent leurs valeurs initiales qui sont celles des nuages. Ce bruit résiduel est susceptible de perturber le processus d'apprentissage par la suite. D'autre part, les polygones de référence des classes pures ne sont parfois pas parfaitement homogènes. Par définition, pour ces classes, les polygones issus de la BD Forêt

sont composés d'au moins 75% d'essences pures et donc les 25% restants peuvent correspondre à d'autres formations végétales, du sol nu ou de l'urbain au sein du même polygone, ce qui peut se traduire par des valeurs de réflectance différentes de celle de l'essence principale. Par ailleurs, et notamment pour les six classes non validées par photo-interprétation (contrairement aux peupleraies), il est possible qu'un polygone ait subi un changement dû à une coupe, des dégâts de tempête, etc. et il peut donc présenter des valeurs de réflectance qui s'écartent de celles des polygones de la même classe.

Afin de limiter la présence des valeurs aberrantes de réflectance au sein des échantillons de référence, une phase de nettoyage des données a été appliquée.

La méthode retenue pour la détection des valeurs est très simple et s'appuie sur la règle des 1,5 Écart Interquartile (EI) qui considère les valeurs situées en dehors de l'intervalle limitée par la frontière basse $Q1^6 - 1.5 \times (Q3 - Q1)$ et la frontière haute $Q3^7 + 1.5 \times (Q3 - Q1)^8$ comme des valeurs aberrantes (Figure 2.9).

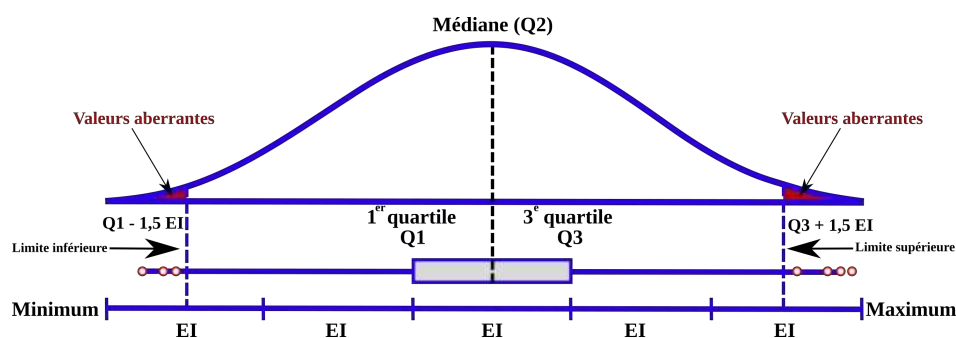


FIGURE 2.9 – Illustration de la règle des $1,5 \times EI$. Elle considère toute valeur en dehors de l'intervalle $[Q1 - 1,5 \times EI, Q3 + 1,5 \times EI]$ comme une valeur aberrante. Q1 et Q3 étant le premier et le troisième quartile.

Dans le cas d'une distribution normale, la zone non couverte par la boîte à moustache qui comprend les valeurs atypiques (*outliers*) représente 0,7% des observations.

La méthode de détection des valeurs aberrantes appliquée consiste d'abord à éliminer tous les pixels situés en dehors de l'intervalle prédéfini pour chaque bande de la série temporelle et ce, pour chacune des sept classes séparément. Ensuite, et en procédant par classe, seuls les pixels valides communs à toutes les bandes de la série temporelle sont retenus. De cette façon, si des pixels sont détectés comme étant aberrants dans une seule bande, ils sont automatiquement retirés de l'ensemble des échantillons.

Pour chacune des trois tuiles de référence, le nombre de pixels retenus à l'issue de cette étape est récapitulé dans le Tableau 2.2 pour chaque année et pour chacune des classes d'essence.

6. Premier quartile (Q1) : valeur supérieure ou égale à au moins 25% des données

7. Troisième quartile (Q3) : valeur supérieure ou égale à au moins 75% des données

8. $EI = Q3 - Q1$

TABLEAU 2.2 – Nombre total des échantillons de référence par classe extraits pour chaque tuile Sentinel-2 après l'étape d'élimination des valeurs aberrantes.

| Année | Nombre de variables extraites | Tuile | Nombre d'échantillons par classe (en pixels) | | | | | | |
|-------|-------------------------------|-------|--|------------------|------------------|-------|------------------|------------------------------------|-------------------------------------|
| | | | Peuplier | Robinier | Châtaignier | Chêne | Hêtre | Forêt fermée de feuillus (mélange) | Forêt ouverte de feuillus (mélange) |
| 2017 | 340 | 31UEQ | 2500 | 2500 | s/o ¹ | 2500 | 2500 | 2500 | 2500 |
| | | 30TYT | 4000 | 4000 | 4000 | 4000 | s/o ¹ | 4000 | 4000 |
| | | 31TCJ | 7700 | 7700 | 7700 | 7700 | NA ¹ | 7700 | 7700 |
| 2018 | 360 | 31UEQ | 3700 | s/o ¹ | s/o ¹ | 3700 | 3700 | 3700 | 3700 |
| | | 30TYT | 3200 | 3200 | 3200 | 3200 | s/o ¹ | 3200 | 3200 |
| | | 31TCJ | 5000 | 5000 | 5000 | 5000 | s/o ¹ | 5000 | 5000 |

¹ Classe non disponible dans la tuile d'étude ou très peu représentée.

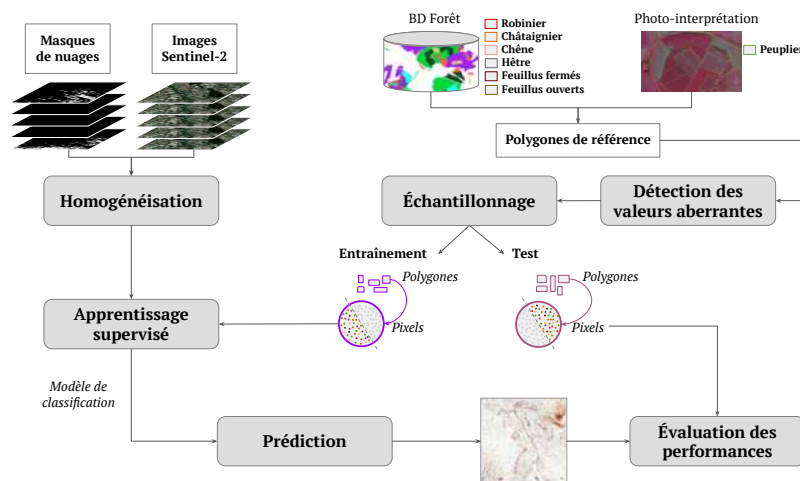
Afin d'assurer un équilibre entre les différentes classes et d'éviter qu'une classe soit sur-représentée ou sous-représentée, nous avons choisi une stratégie d'échantillonnage qui garantit le même nombre de pixels par classe appartenant à la même tuile de la même année. Ce nombre a été fixé de manière à ce que la plus petite classe soit entièrement échantillonnée. Comme cela peut être déduit du Tableau 2.2, la différence entre le nombre d'échantillons en 2017 et 2018 est due au fait qu'il y avait plus de valeurs aberrantes éliminées dans l'année 2018. Pour cette année, la classe des robiniers a été retirée de la tuile 31UEQ car la majorité de ses pixels ont été identifiés comme aberrants. Après vérification, sur les cinq parcelles de robinier disponibles, trois d'entre elles correspondaient à des nuages mal-détectés dans certaines dates. Le nombre de pixels restants était trop faible pour être considéré pour chacune des classes de la tuile.

Chapitre 3

Classification supervisée à partir de séries temporelles d'images satellitaires

Sommaire

| | | |
|-------|---|----|
| 3.1 | Apprentissage automatique | 56 |
| 3.1.1 | Apprentissage non-supervisé | 56 |
| 3.1.2 | Apprentissage supervisé | 57 |
| 3.2 | Choix de l'algorithme de classification | 61 |
| 3.2.1 | Les forêts aléatoires : principe | 62 |
| 3.2.2 | Optimisation des hyper-paramètres | 63 |
| 3.3 | Évaluation des performances | 64 |
| 3.3.1 | Matrice de confusion | 65 |
| 3.3.2 | Mesures de performance | 65 |



3.1 Apprentissage automatique

Le début de l'Intelligence Artificielle (IA) remonte aux années 1950 avec Alan Turing (Turing, 1950) qui cherchait à simuler l'intelligence humaine dans des machines conçues pour « penser » comme les humains et capables de mimer leurs actions. L'IA regroupe donc l'ensemble des techniques qui permettent aux machines d'imiter une forme d'intelligence réelle. L'apprentissage automatique ou *machine learning* est une branche de l'IA. Il est apparu en 1959 avec le programme d'Arthur Samuel (Samuel, 1959) qui jouait au jeu de dames et s'améliorait à chaque partie en mémorisant chaque position jouée et toutes les chances qu'elle offrait. En effet, l'apprentissage automatique permet aux machines d'apprendre à partir d'un ensemble de données et d'expériences sans avoir été explicitement programmées pour le faire. Cela constitue la base d'algorithmes émergents de plus en plus complexes, creusant en profondeur pour générer différentes couches non structurées et se nourrissant d'une quantité considérable de données, on parle ainsi d'une sous-branche de l'apprentissage automatique qui est l'apprentissage profond (Figure 3.1).

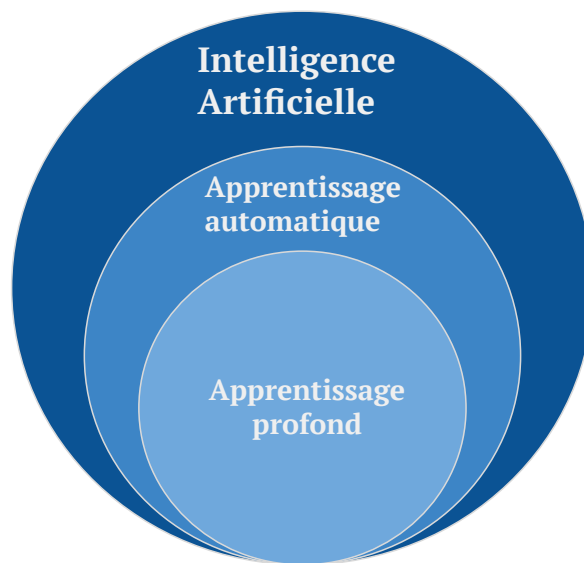


FIGURE 3.1 – L'Intelligence Artificielle et ses dérivés.

Les techniques d'apprentissage automatique sont de plus en plus utilisées en télédétection (Lary et al., 2016), principalement pour la production de cartes thématiques à partir de procédures de classification d'images. En télédétection, il existe deux principaux modes d'apprentissage automatique : supervisé et non supervisé.

3.1.1 Apprentissage non-supervisé

Ce type d'apprentissage consiste à déterminer des groupes similaires ou *clusters* à partir d'un certain nombre de variables ou de caractéristiques sans *a priori* sur leurs classes

d'appartenance, appelées étiquettes ou labels. À partir des différents pixels de l'image, l'algorithme forme d'abord des groupes homogènes liés par des caractéristiques communes. Une étiquette, qui peut être une classe d'occupation du sol (ou toute autre interprétation qui fait sens pour l'expert), est ensuite associée à chacun des clusters (Figure 3.2). Comme évoqué dans la Section 1.3.2, les deux principales contraintes avec les approches de *clustering* sont d'abord l'interprétation des groupes formés qui peut être difficile si aucune expertise ou connaissance préalable n'est disponible pour guider leur formation. Cela nécessite souvent des post-traitements d'agrégation ou de division de *clusters* afin d'obtenir des informations plus cohérentes avec les regroupements thématiques attendus (Wemmert et al., 2009). La deuxième contrainte réside dans le temps de calcul qui est d'autant plus complexe que le nombre des caractéristiques est important (Chen and Gong, 2013).

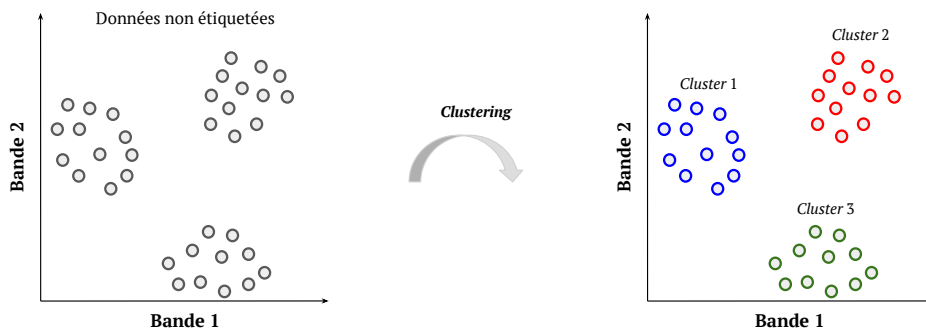


FIGURE 3.2 – Exemple simple d'un *clustering* dans un espace à deux dimensions : bande 1 et bande 2.

Les approches non supervisées sont souvent appliquées afin de fournir une analyse préliminaire permettant de comprendre l'image, ou encore quand peu de données de référence sont disponibles. K-means (Tou et al., 1974) est l'un des algorithmes de *clustering* les plus utilisés en télédétection.

3.1.2 Apprentissage supervisé

L'apprentissage supervisé est guidé par des échantillons de références qui apportent les connaissances préalables sur les étiquettes des classes attendues. En fonction du problème d'apprentissage automatique rencontré, deux approches sont possibles :

- **La régression** : c'est quand il s'agit d'estimer une variable quantitative et continue (variable d'intérêt) en fonction des caractéristiques en entrée (variables explicatives). Les modèles de régression sont par exemple utilisés pour estimer la biomasse à partir de données RADAR (Le Toan et al., 2004).
- **La classification** : elle consiste à prédire une variable qualitative discrète en divisant le jeu de données d'entrée en classes séparables à partir d'une base de références pré-classées. La classification supervisée est l'approche la plus couramment adoptée pour la

production de cartes thématiques à partir d'images satellitaires (Richards, 2013; Gómez et al., 2016). Les algorithmes de classification comprennent des approches simplifiées appelées approches paramétriques et des approches non-paramétriques plus complexes mais plus modulables.

Approches paramétriques et non-paramétriques

Le problème de classification peut se résumer par la définition d'une fonction d'apprentissage : $f : X \rightarrow Y$ permettant de désigner une classe Y à partir des caractéristiques X en entrée. La fonction d'apprentissage peut être simplifiée à une forme connue (par exemple, une fonction linéaire) caractérisée par un ensemble fixe de paramètres quelle que soit la taille de l'échantillon : on parle ainsi d'algorithmes paramétriques. Comme décrit par Richards (2013), « *Quelle que soit la quantité de données que vous soumettez à un modèle paramétrique, il ne changera pas d'avis sur le nombre de paramètres dont il a besoin* ». Ces modèles vont donc apprendre les paramètres de la fonction définie à partir des échantillons de référence. L'Analyse Discriminante Linéaire et la Régression Logistique sont deux exemples d'algorithmes paramétriques largement utilisés. Bien qu'ils soient simples à comprendre et peu exigeants en termes de nombre d'échantillons d'entraînement, les modèles paramétriques sont contraints à respecter une forme spécifique de la fonction d'apprentissage avec un vecteur de paramètres donné simplifiant ainsi le problème d'apprentissage. Ils sont par conséquent plus appropriés pour des problèmes peu complexes sans exiger un nombre important d'échantillons de référence.

Cependant, dès que la quantité des données de référence est importante et que le problème ne peut pas être simplifié dans une fonction d'apprentissage connue, il vaut mieux laisser les données « parler » d'elles-mêmes pour définir la fonction qui s'y ajuste le mieux, ce qui implique l'utilisation d'approches non-paramétriques. Il s'agit d'algorithmes qui ne reposent pas sur des hypothèses strictes quant à la fonction sous-jacente, qui ne sont pas caractérisés par un nombre prédéterminé de paramètres et qui cherchent à s'adapter aux données d'entraînement disponibles. Plus le nombre d'échantillons de référence est élevé plus la quantité d'information exploitable des données est importante. Parmi les algorithmes non-paramétriques les plus fréquemment utilisés dans la communauté de la télédétection figurent les Machines à Vecteurs de Support ou *Support Vector Machine* (SVM) et les Forêts Aléatoires ou *Random Forest* (RF).

Qu'elle soit paramétrique ou non, la tâche d'apprentissage supervisé est au cœur de la procédure de production de cartes thématiques qui sera détaillée dans la section suivante.

Principe de la classification supervisée

La Figure 3.3 schématise tout le processus de classification supervisée à partir d'une série temporelle Sentinel-2, en partant du pré-traitement des entrées, détaillé dans les Sections 2.2.3 et 2.4, jusqu'à l'évaluation des performances.

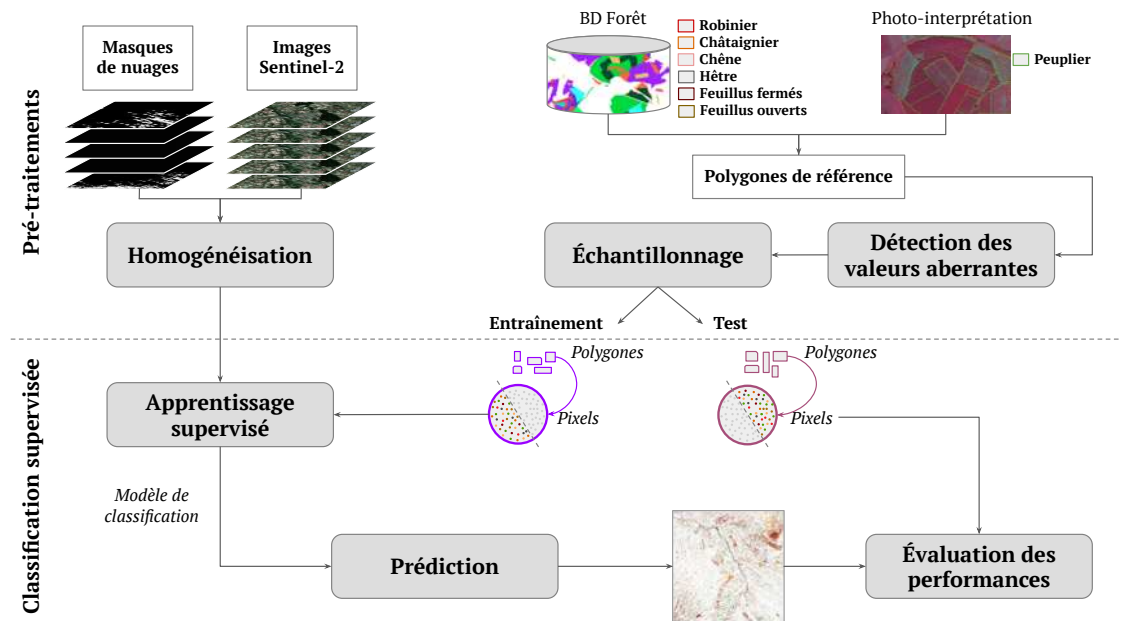


FIGURE 3.3 – Processus de la classification supervisée.

À l'issue des étapes de pré-traitements, les valeurs des pixels sont extraites pour les classes d'intérêt à partir des séries temporelles Sentinel-2 harmonisées. L'espace des caractéristiques résultant est formé, pour chaque pixel, d'un ensemble de variables spectrales (les 10 bandes spectrales aux différentes dates) ainsi qu'une classe associée. Les trois principales étapes de la classification supervisée commencent ensuite, à savoir l'apprentissage d'un modèle, la prédiction (utilisation du modèle appris pour classer des pixels dans l'image) et l'évaluation des performances.

À partir des échantillons d'entraînement, l'algorithme apprend un modèle de classification selon une fonction d'apprentissage. Les échantillons utilisés dans cette étape d'apprentissage doivent être présents en nombre suffisant et représentatifs de la variabilité des classes dans l'image d'étude. En dépit de leur grande dimension, les caractéristiques spectrales qui les décrivent doivent idéalement être similaires au sein d'une même classe (faible variabilité intra-classe) et suffisamment différentes entre les autres classes (forte variabilité inter-classes) afin de mieux les séparer. Bien que réduit lors de l'étape de détection des valeurs aberrantes, la présence du bruit dans les valeurs spectrales (introduit lors de l'acquisition ou des pré-traitements) ou dans les données de référence (échantillons mal identifiés ou contradictoires) est fréquente et ne doit pas être négligée.

D'autre part, la fonction d'apprentissage doit parfaitement s'ajuster aux échantillons d'entraînement, sans être trop contrainte par des hypothèses strictes liées à une forme spécifique de fonction qui synthétise les données par un vecteur de paramètres (pouvant conduire à un sous-apprentissage ou *underfitting*), ni trop spécifique aux échantillons d'entraînement en épousant scrupuleusement la forme de leur fonction sous-jacente au point de les apprendre par cœur (pouvant conduire à un sur-apprentissage ou *overfitting*). Dans la communauté de l'apprentissage automatique, ce problème est connu sous le nom de « dilemme biais-variance », où le biais est lié à l'erreur commise sur les hypothèses de la fonction d'apprentissage et se réfère à la qualité de l'ajustement du modèle aux échantillons d'entraînement, tandis que la variance reflète la dispersion des prédictions autour de leurs valeurs réelles et donc la capacité de généralisation du modèle. Plus le modèle est complexe, plus la variance est importante et plus le modèle est simple, plus le biais est important.

La Figure 3.4 illustre un exemple d'un même problème de classification (deux classes bleue et verte) avec trois fonctions d'apprentissage (dessinées en rouge). La première fonction (Figure 3.4a) présente une situation de sous-apprentissage où la fonction d'apprentissage est trop simple et peu adaptée pour expliquer la variabilité des deux classes. Cela se traduit par un biais élevé et des erreurs de classification de plusieurs échantillons. Contrairement, l'exemple de la Figure 3.4c montre une fonction d'apprentissage avec une forte variance où elle suit parfaitement la forme des données. Bien que tous les échantillons sont correctement classés, l'ajout ou la suppression d'un seul échantillon risquerait de modifier la fonction diminuant ainsi sa capacité de généralisation. Enfin la fonction du milieu (Figure 3.4b) qui présente un bon équilibre entre le biais et la variance.

Il est donc nécessaire de trouver un compromis réalisable entre les situations de sous-apprentissage et de sur-apprentissage, qui permette à la fonction d'apprentissage de décrire efficacement les échantillons d'entraînement avec toute leur dimension spectrale et temporelle, tout en conservant une capacité de généralisation afin de prédire correctement de nouveaux échantillons.

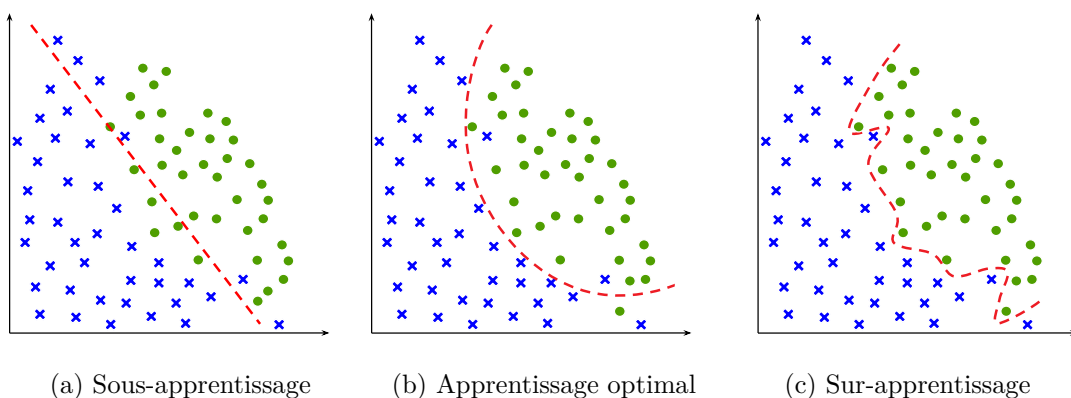


FIGURE 3.4 – Illustration du dilemme biais-variance.

Le modèle de classification résultant est ensuite mobilisé pour prédire toute l'image d'étude. Cette étape consiste à attribuer à chaque pixel de l'image une classe selon les règles de décision apprises lors de l'étape précédente.

Afin d'évaluer la qualité de la prédiction, les échantillons test, qui n'ont pas été exploités jusqu'à présent, sont utilisés pour comparer leurs classes réelles avec celles prédites par le modèle afin d'en déduire des mesures de performance de la classification. Lorsque ces échantillons sont indépendants de ceux utilisés lors de l'apprentissage, l'évaluation est plus fiable.

Comment choisir donc un algorithme qui soit capable de généraliser l'apprentissage sur les données d'entraînement sans être trop influencé par le bruit dans l'espace des caractéristiques, de gérer une grande dimension spectrale et temporelle telle que celle issue des séries temporelles Sentinel-2 tout en assurant de bonnes performances prédictives ?

3.2 Choix de l'algorithme de classification

Pour la classification sur de grandes étendues, le choix de l'algorithme est primordial. Outre les critères abordés dans la section précédente, notamment la capacité de généralisation du classifieur, sa robustesse au bruit et la gestion de données en grande dimension (nombre de variables élevé), il est nécessaire de prendre en compte d'autres facteurs afin de limiter les contraintes de la large échelle. Il s'agit d'abord du temps de calcul qui doit être raisonnable principalement pour les étapes les plus chronophages d'apprentissage et de prédiction. D'autre part, l'algorithme choisi doit idéalement offrir des performances dont l'influence est limitée par la valeur des hyper-paramètres afin de lancer des classifications sur de grandes étendues à différents moments sans avoir à redéfinir ou optimiser à nouveau les paramètres. Par ailleurs, le classificateur ne doit pas être très exigeant en termes de nombre d'échantillons d'entraînement et sans que cela ait un impact direct sur les performances.

Les travaux récents avec les séries temporelles d'images satellitaires ont rapporté l'efficacité de l'algorithme des forêts aléatoires (RF) dans le cadre d'approches opérationnelles sur de grandes étendues (Mellor et al., 2013; Balzter et al., 2015; Pelletier et al., 2016; Inglada et al., 2017). Certains auteurs (Meyer et al., 2016; Pelletier et al., 2016) ont comparé l'efficacité du RF par rapport à SVM et ont montré que SVM requiert un temps d'apprentissage beaucoup plus long que celui du RF avec des performances généralement plus faibles. D'autres ont montré la robustesse du RF face à la présence de bruit dans les données d'entraînement (Breiman, 2001; Cutler et al., 2007) avec une plus grande stabilité en comparaison avec SVM (Pelletier et al., 2017). Par ailleurs, bien que certains travaux (Dalponte et al., 2013; Ghosh et al., 2014) aient considéré que les deux algorithmes étaient aussi fiables l'un que l'autre, RF s'est avéré plus efficace lorsqu'il s'agissait de données de grande dimension (Chan et al., 2012; Pelletier et al., 2016). Cependant, pour éviter le sur-apprentissage avec le RF, il est souvent

recommandé d'optimiser ses hyper-paramètres (Segal, 2004).

Tous ces avantages, y compris le temps de calcul rapide, l'efficacité face à la grande dimension des données ainsi que les bonnes performances avec une taille d'échantillons suffisante, ont fait du RF un bon candidat pour la classification des peupleraies sur de grandes étendues.

3.2.1 Les forêts aléatoires : principe

RF (Breiman, 2001) est un algorithme de classification non-paramétrique. Il s'agit d'un modèle d'apprentissage ensembliste qui utilise des arbres de décision comme classifieur de base. Le principe des méthodes ensembliste est de combiner plusieurs classifieurs, souvent peu performants individuellement, afin d'obtenir un résultat plus performant qu'avec un classifieur unique (Rodriguez-Galiano et al., 2012). Ceci résulte en un modèle plus stable (variance atténuée) et plus flexible (biais réduit). En effet, RF est basé sur la méthode de *bagging* (de *bootstrap aggregating*) qui consiste à former plusieurs arbres décisionnels¹, où chaque arbre est entraîné avec un sous-ensemble aléatoire et différent d'échantillons d'entraînement appelés échantillons *bootstrap* (Breiman, 2001). Chaque sous-ensemble d'échantillons *bootstrap* est obtenu par tirage aléatoire avec remise de n échantillons de référence permettant de construire les t arbres de décision.

Chaque arbre de décision commence par une racine commune qui est étendue au fur et à mesure en formant des nœuds. Au niveau de chaque nœud « parent », un ensemble de variables m (les caractéristiques qui décrivent les échantillons) est sélectionné aléatoirement et sans remise pour séparer les échantillons d'entraînement en deux nœuds « enfants » plus homogènes. En général, $m = \sqrt{f}$ où f représente le nombre total de variables. La meilleure séparation est ensuite déterminée en se basant sur une mesure d'impureté qui est généralement² l'indice de Gini. Plus sa valeur est faible plus le nœud est pur. La décision d'arrêter l'expansion du nœud revient à la formation d'une feuille. Un nœud feuille est un nœud pur comportant des échantillons homogènes qui sont affectés à la même classe. Le modèle final est ainsi construit par *bagging*, où les t prédicteurs indépendants sont agrégés par vote majoritaire. Les arbres sont faiblement corrélés et, bien que leurs performances individuelles soient faibles, elles sont nettement meilleures lorsqu'ils sont regroupés.

Pour la prédiction de nouveaux échantillons, chaque arbre vote pour une classe donnée et c'est la classe majoritaire qui est retenue. La combinaison des prédictions issues des arbres de décision individuels appris sur des échantillons différents permet de diminuer la variance, augmentant ainsi la capacité de généralisation du RF.

Trois principaux hyper-paramètres peuvent être définis pour le RF, à savoir le nombre

1. Un arbre de décision consiste en une suite de décisions séquentielles et binaires permettant de prédire une variable d'intérêt quantitative ou qualitative.

2. L'entropie peut également être considérée comme une mesure d'impureté du nœud.

d'arbres, la profondeur maximale de chaque arbre et le nombre de variables aléatoires sélectionnées dans chaque nœud. D'autres hyper-paramètres sont plus rarement considérés, notamment le nombre minimal d'échantillons dans chaque nœud intermédiaire ou pur, la mesure d'impureté et le nombre maximal d'échantillons par arbre.

3.2.2 Optimisation des hyper-paramètres

Hyper-paramètres

Bien qu'il existe des recommandations dans la littérature pour définir les valeurs de certains hyper-paramètres, il est difficile de les déterminer *a priori* et de conserver les mêmes valeurs pour des jeux de données différents. Il est donc nécessaire d'optimiser les hyper-paramètres en fonction du jeu de données concerné afin de tendre vers les meilleures performances de classification.

En ce qui concerne le nombre d'arbres, il n'y pas de consensus clair pour le déterminer. L'augmentation du nombre d'arbres peut paraître intuitivement synonyme d'amélioration des performances. Cependant, il a été démontré que pour chaque jeu de données, il existe un nombre optimal d'arbres, mais l'augmentation de ce nombre n'implique pas nécessairement une augmentation des performances mais plutôt du coût de calcul (Oshiro et al., 2012). Dans un article plus récent de Probst and Boulesteix (2018), les auteurs ne recommandent pas une optimisation du nombre d'arbres dans le cas de la classification et indiquent que le gain en performance est généralement atteint avec les 100 premiers arbres. À notre connaissance, il n'existe pas de travaux qui fournissent des recommandations pour le choix du nombre d'arbre en fonction du nombre de variables d'entrée ou du nombre d'échantillons. On peut toutefois supposer que le nombre d'arbres ne doit pas être trop important par rapport au nombre de variables, afin d'éviter de générer des arbres individuels trop corrélés, ce qui n'est pas le but de l'algorithme.

La profondeur maximale des arbres individuels a également tendance à avoir peu d'impact sur la performance de classification (Segal, 2004). Mais comme le recommande Hastie et al. (2009), étant donné que l'augmentation de la profondeur des arbres n'est généralement pas très coûteuse en termes de temps de calcul, il est préférable de la fixer à sa valeur maximale afin d'avoir un hyper-paramètre de moins à optimiser.

Le nombre de variables aléatoires à sélectionner pour séparer chaque nœud a été davantage étudié dans la littérature. Il semble avoir une influence significativement plus importante sur les performances que les deux hyper-paramètres précédents. En effet, lorsque le nombre de caractéristiques disponibles est important, mais que la proportion de celles qui sont pertinentes est faible, RF risque d'être peu performant avec un nombre réduit de variables (Hastie et al., 2009). Bien que RF soit robuste à la grande dimension des données et au sur-apprentissage, il

est souvent préconisé d'éviter d'avoir un très grand nombre de variables comparé au nombre d'échantillons, voire même d'éliminer les variables peu pertinentes pour garantir les meilleures performances (Belgiu and Drăguț, 2016). La valeur recommandée pour le nombre de variables permettant de diviser un nœud est égale à la racine carrée du nombre total des caractéristiques (Ghosh and Joshi, 2014; Belgiu and Drăguț, 2016).

Optimisation par validation croisée

L'optimisation des hyper-paramètres du RF consiste à trouver les paramètres optimaux qui donnent les meilleures performances de classification. L'approche adoptée est une approche par grille de recherche, qui consiste à définir une liste de valeurs pour chaque hyper-paramètre, d'évaluer chaque combinaison possible dans un modèle afin de déterminer celle qui garantit les prédictions les plus précises.

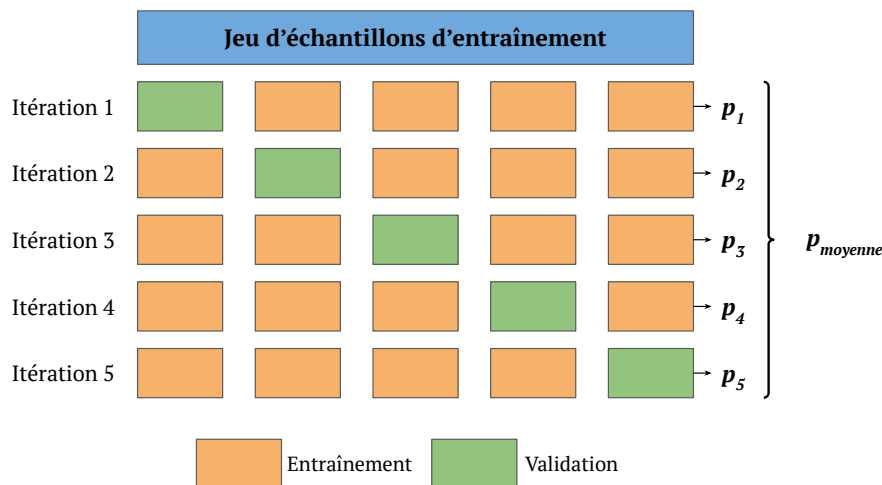


FIGURE 3.5 – Illustration du principe de la validation croisée k -fold avec $k = 5$.

L'évaluation est effectuée à partir du jeu d'entraînement avec une approche de validation croisée k -fold et selon une valeur de performance p prédéfinie (Section 3.3). En effet, pour chaque combinaison candidate, les échantillons sont divisés en k sous-ensembles différents où chaque $k - 1$ sont utilisés pour l'**entraînement** du modèle qui sera ensuite évalué et **validé** sur chaque sous-ensemble restant. k évaluations en résultent pour une même combinaison et c'est la valeur de performance moyenne qui est considérée. Le principe de la validation croisée est illustré dans la Figure 3.5 pour $k = 5$.

3.3 Évaluation des performances

L'évaluation de la qualité des classifications est effectuée à partir du jeu d'échantillons test n'ayant pas servi à l'apprentissage. Cette étape consiste à quantifier les écarts entre les classes réelles des échantillons test et celles prédites par le classifieur. Ces différences sont

généralement synthétisées dans une matrice appelée matrice de confusion, à partir de laquelle diverses mesures de performance peuvent être dérivées.

Il est important de noter que deux modèles sont construits : un premier modèle complet qui est élaboré avec tous les échantillons de référence (entraînement et test) et qui sert à prédire et à construire le produit cartographique, et un modèle qui s'appuie sur un sous-ensemble des échantillons et qui est uniquement destiné à évaluer les performances du modèle complet.

3.3.1 Matrice de confusion

La matrice de confusion est une matrice carrée de taille $n \times n$ où n est le nombre de classes. Les lignes i représentent généralement les classes réelles alors que les colonnes j indiquent les classes prédites.

$$\begin{bmatrix} c_{11} & \dots & c_{1n} \\ \vdots & \ddots & c_{ij} \\ c_{n1} & & c_{nn} \end{bmatrix}$$

Le coefficient c_{ij} représente le nombre d'échantillons test de la classe i prédits par le classifieur dans la classe j . Une classification sans erreur se traduit par une matrice diagonale où chaque coefficient c_{ii} de la diagonale principale est égal au nombre d'échantillons test de la classe i et les coefficients c_{ij} extérieurs sont nuls.

3.3.2 Mesures de performance

À partir de la matrice de confusion, des métriques de performance sont calculées afin d'évaluer la qualité des prédictions de manière globale ou par classe.

Métriques globales

Deux indices sont couramment utilisés pour évaluer la classification de manière globale : la précision globale (*Overall Accuracy* (OA)) et l'indice *Kappa*. L'OA reflète la proportion des bonnes classifications et elle est exprimée par le nombre d'échantillons correctement prédits (la diagonale de la matrice de confusion) divisé par le nombre total d'échantillons test (soit le taux d'accord global) :

$$OA = \sum_{i=1}^n c_{ii} \frac{1}{N} \quad (3.1)$$

où le nombre d'échantillons test $N = \sum_{i=1}^n \sum_{j=1}^n (c_{ij})$.

Bien que son interprétation soit assez intuitive, l'OA seule ne fournit pas d'information sur les classes correctement ou mal prédites, elle en fait la synthèse de manière globale.

L'indice *Kappa* est également très utilisé en télédétection. Il correspond à une version pénalisée de l'OA qui est « corrigée » de l'effet des bonnes classifications attribuées au hasard. Le *Kappa* mesure donc le taux d'accord après avoir éliminé la part obtenue par hasard en réalisant une classification aléatoire. Le *Kappa* est formulé de la manière suivante :

$$Kappa = \frac{OA - e}{1 - e} \quad (3.2)$$

où e représente la pénalité due au hasard : $e = \frac{1}{N^2} \sum_{i=1}^n (\sum_{j=1}^n c_{ij})(\sum_{j=1}^n c_{ji})$.

Il est de plus en plus recommandé d'abandonner l'utilisation du *Kappa* (Olofsson et al., 2014; Foody, 2020) en raison de plusieurs biais qui sont notamment liés à sa formulation (Foody, 2002) et à son incohérence ainsi qu'à sa redondance avec l'OA. En effet, le *Kappa* est basé sur les totaux en lignes et en colonnes (effectifs marginaux) de la matrice de confusion (intégrés dans le terme e de l'équation (3.2)) qui ne peuvent pas être aléatoires pourtant c'est ce que l'on considère en les retirant de OA. En outre, le calcul du *Kappa* ne prend pas en compte tous les éléments de la matrice de confusion, seuls les accords et les effectifs marginaux sont considérés mais pas les désaccords. Par conséquent, deux valeurs de *kappa* identiques peuvent ne pas refléter les mêmes confusions entre classes, ce qui n'est pas attendu d'un indicateur de qualité. Pour toutes ces limitations et bien d'autres, l'indice *Kappa* a été délibérément écarté de ces travaux de thèse.

Afin de mieux comprendre ce qui sous-tend ces métriques globales, il est nécessaire de déduire de la matrice de confusion des mesures de performance spécifiques à chaque classe.

Métriques par classe

Trois mesures de qualité par classe sont couramment utilisées : la précision utilisateur ou *User's Accuracy* (UA) appelée aussi *précision*, la précision producteur ou *Producer's Accuracy* (PA) dénommée *rappel*, et leur moyenne harmonique, le *F-score*.

- UA : pour une classe i , elle correspond à la proportion des échantillons correctement prédits parmi toutes les prédictions observées pour cette même classe. Elle est formulée de la manière suivante :

$$UA_i = \frac{c_{ii}}{\sum_{j=1}^n c_{ji}} \quad (3.3)$$

- PA : pour une classe i , elle représente la proportion des échantillons correctement prédits parmi tous les échantillons appartenant à la classe. Elle est formulée comme suit :

$$PA_i = \frac{c_{ii}}{\sum_{j=1}^n c_{ij}} \quad (3.4)$$

- F-score : c'est une combinaison des deux mesures précédentes. Pour une classe i , elle

correspond à la moyenne harmonique des UA et PA :

$$F - score_i = 2 \frac{UA \times PA}{(UA + PA)} \quad (3.5)$$

Il est important de noter que la plupart de ces métriques de performance sont sensibles à la prévalence, qui traduit le nombre d'échantillons disponibles par classe. Un déséquilibre dans la représentativité des différentes classes peut avoir un impact sur l'évaluation des performances de classification (Foody, 2020). Afin de réduire l'influence de la prévalence, il est préconisé de privilégier des ensembles d'échantillons équilibrés et d'utiliser des mesures indépendantes de la prévalence telle que le PA par classe. En outre, afin de rendre l'estimation de la qualité plus fidèle à la réalité, il est souvent recommandé d'utiliser une mesure globale, principalement l'OA, conjointement avec des métriques par classe, notamment les UA et PA (Olofsson et al., 2014).

Dans le cadre de ces travaux de thèse, l'OA a été utilisée comme une mesure synthétique de la performance de classification ainsi que les trois métriques par classe les UA_i , PA_i et $F-score_i$.

Tous les traitements de classification, d'optimisation des hyper-paramètres et d'évaluation des performances ont été élaborés en Python avec la bibliothèque *scikit-learn* (Pedregosa et al., 2011).

Troisième partie

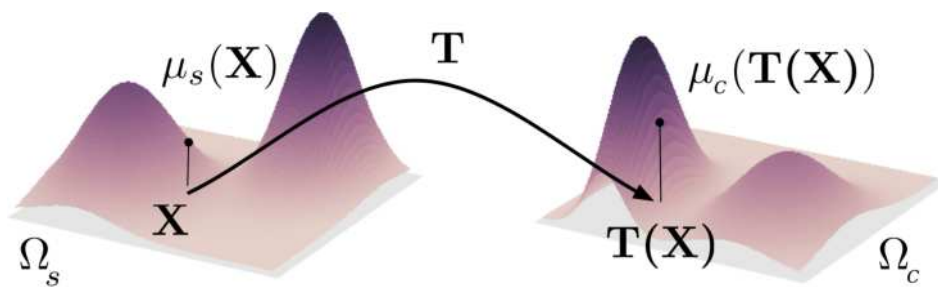
**Adaptation de domaine pour la
classification à large échelle des peupleraies**

Chapitre 4

Adaptation de domaine par transport optimal

Sommaire

| | | |
|-------|--|-----|
| 4.1 | Principe du transport optimal | 70 |
| 4.1.1 | La formulation initiale de Monge | 70 |
| 4.1.2 | Formulation de Kantorovich : formulation continue et infinie | 72 |
| 4.1.3 | Formulation de Monge-Kantorovich | 73 |
| 4.1.4 | Adaptation de domaine par transport optimal | 74 |
| 4.2 | Application du transport optimal pour l'identification des peupleraies sur de grandes étendues | 80 |
| 4.2.1 | Protocole expérimental | 81 |
| 4.2.2 | Transfert spatial | 86 |
| 4.2.3 | Transfert temporel | 97 |
| 4.2.4 | Transfert spatio-temporel | 101 |
| 4.3 | Conclusion | 107 |
| 4.4 | Bilan | 109 |



Dans ce chapitre, nous explorons une technique non supervisée d'adaptation de domaine par adaptation des distributions de données - le transport optimal - afin d'évaluer son potentiel pour la classification automatique des peupleraies à partir d'images sur de grandes étendues. Nous introduisons d'abord les concepts de base du transport optimal à travers un exemple théorique permettant de mieux appréhender les notions. Nous présentons ensuite une application de différentes variantes du transport optimal sur les séries temporelles Sentinel-2. Enfin, nous évaluons l'intérêt et les limites de l'approche pour une application à large échelle.

4.1 Principe du transport optimal

Le problème de transport optimal a été initialement formalisé par Monge au cours du XVIII^e siècle, puis reformulé par Kantorovich au milieu du XX^e siècle pour aboutir désormais à la formulation dite de Monge-Kantorovich. Il s'agit d'un problème d'optimisation qui consistait à l'origine à trouver le moyen le plus économique de transporter des objets depuis différents points de départ ou sources (S) vers différents points d'arrivée ou cibles (C) (Kantorovich, 2006). Les améliorations apportées par Kantorovich ont permis une meilleure généralisation du problème de transport optimal, ce qui a donné lieu à de nombreuses applications dans les domaines militaire, économique et financier.

Les formulations mathématiques seront introduites à travers un exemple simple de transport de grumes de peupliers depuis diverses peupleraies vers plusieurs industries de transformation de bois. Cet exemple est inspiré de celui de Peyré (2019a).

4.1.1 La formulation initiale de Monge

Considérons une entreprise de transport de bois qui achemine des grumes de peuplier du lieu de leur production (points rouges) jusqu'aux industries de transformation (points bleus) comme illustré dans la Figure 4.1. Dans cet exemple, nous admettons qu'il existe cinq peupleraies et cinq industries de transformation. L'objectif étant de minimiser le coût global du transport de grumes.

Supposons également que chaque peupleraie produit la même quantité de grumes q et que chaque industrie de transformation ne peut accepter que cette quantité q . Ainsi, une industrie de transformation ne peut collecter que les grumes d'une seule peupleraie et cette dernière ne peut être destinée qu'à une seule industrie de transformation, ce qui implique qu'il y a autant de peupleraies (n) que d'industries de transformation (m). Il est donc possible d'exprimer \mathcal{S}_n comme étant le nombre de permutations (s) possibles entre chaque peupleraie i et chaque industrie j :

$$s : i \in \{1, 2, 3, 4, 5\} \mapsto j \in \{1, 2, 3, 4, 5\} \quad (4.1)$$



FIGURE 4.1 – Illustration d’une seule permutation pour un transport optimal entre l’ensemble des peupleraies et des industries de transformation selon la formulation de Monge. Les points rouges représentent les lieux de départ (les peupleraies) et les points bleus correspondent aux points de destination (les industries de transformation).

Il s’agit d’une bijection entre l’ensemble de peupleraies et d’industries, pour laquelle chaque élément i possède une image $s(i) = j$.

L’hypothèse originelle de Monge suppose que le coût de déplacement de chaque unité de masse est égal à la distance parcourue. Dans cet exemple, le coût de déplacement est égal au temps de parcours $C_{i,j}$ entre la peupleraie $i \in \{1, 2, 3, 4, 5\}$ et l’industrie de transformation $j \in \{1, 2, 3, 4, 5\}$. Pour une permutation s , le coût total C est donc égal à la somme des $C_{i,s(i)}$:

$$C = \sum_{i=1}^n C_{i,s(i)} \tag{4.2}$$

Le Tableau 4.1 synthétise l’ensemble des coûts entre chaque peupleraie et chaque industrie de transformation.

TABLEAU 4.1 – Temps de parcours entre les peupleraies (i) et les industries de transformation (j). La permutation s choisie dans la Figure 4.1 est colorée en vert dans le tableau.

| $C_{i,j}$ | $j = 1$ | $j = 2$ | $j = 3$ | $j = 4$ | $j = 5$ |
|-----------|---------|---------|---------|---------|---------|
| $i = 1$ | 41 | 51 | 120 | 140 | 27 |
| $i = 2$ | 67 | 22 | 61 | 67 | 43 |
| $i = 3$ | 90 | 45 | 31 | 47 | 65 |
| $i = 4$ | 27 | 18 | 72 | 106 | 38 |
| $i = 5$ | 68 | 62 | 40 | 103 | 86 |

Le choix de la Figure 4.1 est représenté par les cellules vertes pour lequel :

$$s(1) = 5, s(2) = 2, s(3) = 4, s(4) = 1, s(5) = 3$$

Le coût de transport dans cette configuration (coût de la permutation) où $n = 5$ est donc égal :

$$\begin{aligned} C &= \sum_{i=1}^5 C_{i,s(i)} \\ &= C_{1,5} + C_{2,2} + C_{3,4} + C_{4,1} + C_{5,3} \\ &= 27 + 22 + 47 + 27 + 40 = 163 \end{aligned} \tag{4.3}$$

Selon cette formulation de Monge, le problème de transport optimal est un problème d'optimisation qui vise à trouver la permutation s pour laquelle le coût C est minimal :

$$\min_{s \in \mathcal{S}_n} C(s) \tag{4.4}$$

Comme cela peut être déduit du Tableau 4.1, le nombre de permutations possibles est égal à $n = 5!$, où le point $i = 1$ a $n = 5$ choix possibles, le point $i = 2$ en a $(n - 1) = 4$, et ainsi de suite aboutissant à un total de 120 possibilités.

L'une des principales contraintes avec la formulation de Monge est que le nombre de permutations augmente très rapidement avec n ce qui limite l'application de cette approche aux petites valeurs de n .

4.1.2 Formulation de Kantorovich : formulation continue et infinie

L'amélioration apportée par Kantorovich consiste à modifier le problème de Monge en remplaçant l'ensemble des permutations $s \in \mathcal{S}_n$ par $n!$ matrices de permutation \mathcal{P}_n . Ce sont des matrices de taille $n \times n$ comportant des valeurs binaires où **1** implique une permutation entre i et j et 0 indique une absence de permutation. Dans ce cas, la matrice de coût présentée dans le Tableau 4.1 devient la matrice de permutation suivante :

$$\begin{bmatrix} 0 & 0 & 0 & 0 & \mathbf{1} \\ 0 & \mathbf{1} & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathbf{1} & 0 \\ \mathbf{1} & 0 & 0 & 0 & 0 \\ 0 & 0 & \mathbf{1} & 0 & 0 \end{bmatrix}$$

Le problème de Monge peut donc être réécrit comme suit :

$$\min_{P \in \mathcal{P}_n} \sum_{i=1}^n \sum_{j=1}^n P_{i,j} C_{i,j} \quad (4.5)$$

Jusqu'à ce stade, il s'agit simplement d'une réécriture du problème de Monge. L'apport de Kantorovich se situe ensuite dans la transformation de l'ensemble discret \mathcal{P}_n (grand et fini de $n!$ matrices) en un ensemble continu et potentiellement infini mais plus simple. Ceci consiste à supposer que les entrées de P ne sont pas exclusivement égales à 0 ou à 1 mais des valeurs situées dans l'intervalle $[0, 1]$. Ceci permet de définir un ensemble plus grand de matrices bistochastiques \mathcal{B}_n , qui sont des matrices dont la somme de chaque ligne et de chaque colonne est égale à 1 :

$$\mathcal{B}_n := \left\{ P \in [0, 1]^{n \times n} \mid \forall i, \sum_j P_{i,j} = 1, \forall j, \sum_i P_{i,j} = 1 \right\} \quad (4.6)$$

La formulation du problème de transport optimal devient la suivante :

$$\min_{P \in \mathcal{B}_n} \sum_{i=1}^n \sum_{j=1}^n P_{i,j} C_{i,j} \quad (4.7)$$

L'intérêt de cette reformulation consiste dans l'appartenance de P à l'ensemble \mathcal{B}_n , qui est un ensemble convexe où la moyenne de deux matrices bistochastiques est également bistochastique contrairement aux matrices binaires de permutation (\mathcal{P}_n). La convexité du problème a pu tirer profit du développement de l'algorithme du simplexe à la même époque qui était destiné à la résolution de certains problèmes convexes d'optimisation, dont celui de l'équation (4.7). Comme mentionné par (Peyré, 2019b), pour $n = 20$ un algorithme du simplexe (avec une complexité de n^3) prend environ 0.01 seconde, alors qu'une recherche des $n!$ possibilités durerait 100 000 ans environ. Grâce à ces deux avancées, le champ d'application du transport optimal a pu être élargi à des problèmes de très grande taille.

4.1.3 Formulation de Monge-Kantorovich

La formulation initiale posée par Monge possède une autre contrainte. En effet, l'hypothèse de départ considère qu'il y a autant de points de départ que de points d'arrivée. Si ce nombre est différent alors le problème de transport optimal n'a pas de solution, car il est impossible de mettre en bijection deux espaces de tailles différentes. La généralisation du problème consiste à ne plus considérer les n peuplées et les m industries de transformation où $n = m$, mais plutôt les distributions de quantités de grumes produites (a_1, \dots, a_n) et les capacités des industries de transformation (b_1, \dots, b_m) . Ces quantités doivent être positives et respecter la contrainte de

conservation de la masse, c'est-à-dire qu'il y a autant de grumes que de capacités d'accueil :

$$a_1 + \dots + a_i + \dots + a_n = b_1 + \dots + b_j + \dots + b_m \quad (4.8)$$

Il s'agit ainsi de remplacer les matrices bistochastiques (4.6) par des matrices de couplage où à chaque fois qu'il y a une valeur non nulle dans $P_{i,j}$, il y a un transfert de masse entre i et j :

$$\mathcal{B}_{a,b} := \left\{ P \in \mathbb{R}_+^{n \times n} \mid \forall i, \sum_j P_{i,j} = a_i, \forall j, \sum_i P_{i,j} = b_j \right\} \quad (4.9)$$

Contrairement aux matrices bistochastiques pour lesquelles l'unique solution est une permutation qui consiste à transférer la totalité de la masse a_i d'un élément i vers un élément j , la matrice de couplage P peut avoir plus d'une solution. Dans l'exemple considéré, cela implique que les grumes produites par une peupleraie peuvent être destinées à plusieurs industries de transformation, tout en respectant la contrainte $\sum_j P_{i,j} = a_i$. Ainsi, à chaque fois qu'il y a un transfert de masse, il y a un coût qui est calculé et comme cela demeure une formulation convexe, le problème peut être efficacement résolu par l'algorithme du simplexe.

Cette généralisation de Monge-Kantorovich et les améliorations récentes des algorithmes d'optimisation ont rendu possible l'utilisation du transport optimal dans une multitude d'applications, notamment dans l'adaptation de domaine.

4.1.4 Adaptation de domaine par transport optimal

Principe de base

L'adaptation de domaine par transport optimal a été introduite par [Courty et al. \(2014\)](#). Il s'agit d'une méthode non supervisée d'adaptation de la distribution des données entre un domaine de départ ou source Ω_s et un domaine d'arrivée ou cible Ω_c .

L'objectif est de réduire la distorsion entre les deux domaines ou le *datashift* afin d'adapter les deux distributions à un même classifieur. La Figure 4.2 montre un exemple de deux distributions de points dans un domaine source et cible à deux dimensions et avec 30 échantillons chacun. Les deux domaines comportent trois classes identiques présentant des distributions différentes.

Dans ce cas d'adaptation de domaine, on suppose qu'il y a suffisamment de connaissances sur le domaine source, y compris la distribution et les étiquettes des données, et moins d'informations sur le domaine cible où seules les distributions sont disponibles, ce qui donne lieu à un problème d'adaptation de domaine considéré comme non supervisé. Le problème devient semi-supervisé ou supervisé à partir du moment où certaines étiquettes dans le domaine cible

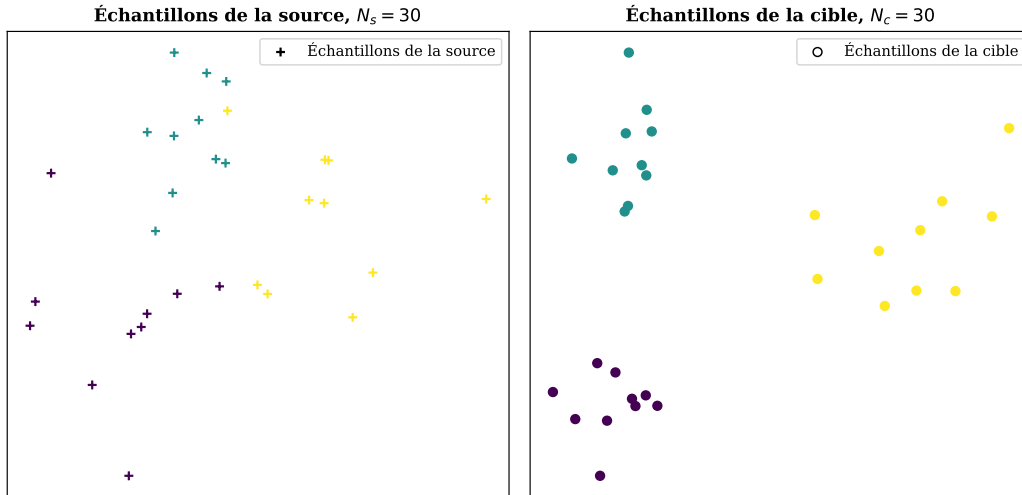


FIGURE 4.2 – Exemple de deux distributions de points dans un espace de caractéristiques pour la source et la cible. Les trois couleurs représentent trois classes différentes. Les étiquettes du domaine cible sont illustrées uniquement à des fins de visualisation.

sont également disponibles (Courty et al., 2014).

La situation peut être modélisée de la manière suivante en considérant :

- $\Omega_s = \{x_i^s, y_i^s\}_{i=1}^{N_s}$ où $x_i^s \in \mathbb{R}^d$ est un échantillon de la source ayant une dimension d (nombre de variables), y_i^s désigne son étiquette (classe) associée et N_s est le nombre d'échantillons disponibles,
- $\Omega_c = \{x_i^c\}_{i=1}^{N_c}$ où $x_i^c \in \mathbb{R}^d$ est un échantillon de la cible ayant une dimension d mais sans information sur son étiquette et N_c représente le nombre d'échantillons disponibles,
- Ω_s et Ω_c sont définis par leurs deux lois de probabilité marginales μ_s et μ_c .

Le postulat de départ suppose qu'il existe une fonction de transformation $T : \Omega_s \rightarrow \Omega_c$ qui transporte μ_s sur μ_c (Figure 4.3).

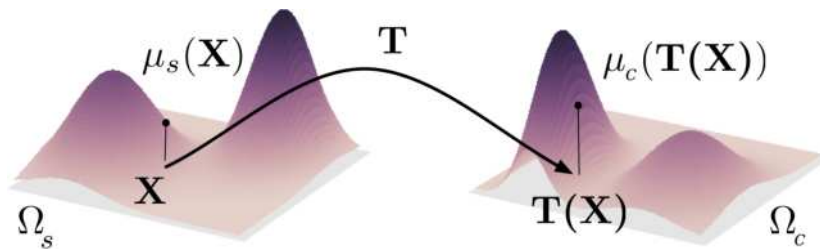


FIGURE 4.3 – Transport optimal des deux mesures de probabilité depuis Ω_s vers Ω_c . Source : <http://people.irisa.fr/Nicolas.Courty/OATMIL/>

Cette fonction doit respecter la distribution conditionnelle des deux domaines, c'est-à-dire que les étiquettes doivent être préservées après la transformation :

$$P_s(y|x^s) = P_c(y|T(x^s)) \quad (4.10)$$

où P_s et P_c désignent les mesures de probabilité sur Ω_s et Ω_c . On cherche donc à trouver l'ensemble \mathcal{B} des couplages $\gamma \in (P_s, P_c)$ entre Ω_s et Ω_c :

$$\mathcal{B} = \left\{ \gamma \in (\mathbb{R}^+)^{N_s \times N_c} \mid \gamma \mathbf{1}_{N_c} = \mu_s, \gamma^T \mathbf{1}_{N_s} = \mu_c \right\} \quad (4.11)$$

où $\mathbf{1}_d$ est un vecteur unitaire de dimension d . La solution du transport optimal qui minimise le coût est γ_0 tel que :

$$\gamma_0 = \underset{\gamma \in \mathcal{B}}{\operatorname{argmin}} \langle \gamma, \mathbf{C} \rangle_F \quad (4.12)$$

avec :

- $C \geq 0$ est la matrice de coût où chaque élément $c(x_i^s, x_j^c)$ représente le coût associé à l'énergie requise pour déplacer une masse de probabilité x_i^s vers x_j^c et qui peut être interprétée comme une mesure de distance,
- $\langle \dots \rangle_F$ désigne le produit scalaire de Frobenius.

La fonction de coût adoptée pour le calcul des couples optimaux est la distance Euclidienne définie comme suit :

$$C(i, j) = \|x_i^s - x_j^c\|^2 \quad (4.13)$$

Comme c'est la distance Euclidienne qui est considérée comme fonction de coût, c'est la distance de Wasserstein connue sous *Earth Mover's Distance* (EMD) qui est utilisée pour évaluer les distances entre les deux distributions de probabilité.

La formulation avec l'EMD de l'équation (4.12) constitue la formulation de départ pour d'autres variantes de transport optimal pour l'adaptation de domaine. Elle constituera la première approche de transport optimal non supervisé envisagée dans la partie applicative de la Section 4.2.

Transport optimal régularisé

La régularisation est une pratique très courante qui permet d'inclure de nouvelles propriétés dans une formulation d'origine (Courty et al., 2014). Dans le cas du transport optimal, elle permet de relaxer certaines contraintes ou d'exploiter d'autres informations disponibles dans la source ou dans la cible (outre les distributions) afin de trouver la fonction de transport la plus appropriée.

- **Régularisation par l'entropie :** Cuturi (2013) a proposé de régulariser le problème

de transport optimal par l'entropie de la matrice de couplage :

$$\begin{aligned}\gamma_0^\lambda &= \operatorname{argmin}_{\gamma \in \mathcal{B}} \langle \gamma, \mathbf{C} \rangle_F + \lambda \Omega_s(\gamma) \\ &= \operatorname{argmin}_{\gamma \in \mathcal{B}} \langle \gamma, \mathbf{C} \rangle_F - \frac{1}{\lambda} h(\gamma)\end{aligned}\quad (4.14)$$

où λ est le terme de régularisation par l'entropie et $h(\gamma)$ est l'entropie de γ .

Comme la majorité des éléments de cette matrice sont nuls avec une forte probabilité, il est possible de diminuer cette dispersion en augmentant leur entropie. L'inclusion du terme de régularisation par l'entropie dans l'équation 4.14 permettra d'avoir moins d'éléments nuls et par conséquent un couplage plus dense entre les distributions. Dans ce cas, un élément de la source pourra avoir plus d'un lien avec les éléments de la cible. Ces liens affecteront la manière dont l'échantillon source est transporté. Plus la valeur de λ est élevée plus le couplage entre les deux domaines est dense. La solution optimale retenue est celle qui a le moins d'éléments nuls et donc la valeur d'entropie la plus élevée. Ce problème peut être résolu avec l'algorithme de Sinkhorn. Cette nouvelle formulation sera référée comme celle de Sinkhorn. Un exemple de deux matrices de couplage avec l'EMD et Sinkhorn est fourni dans la Figure 4.4. Les permutations résultantes sont illustrées dans la Figure 4.5.

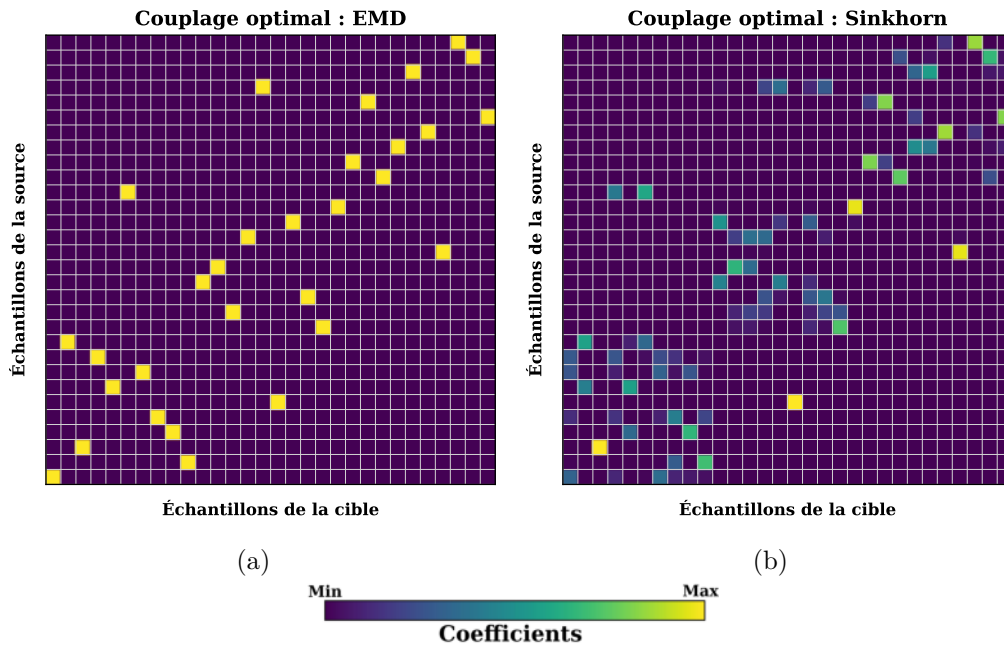


FIGURE 4.4 – Matrices des couples optimaux entre les 30 échantillons de la source représentés par les 30 lignes et les 30 échantillons de la cible représentés par les 30 colonnes avec (a) l'EMD et (b) Sinkhorn. Avec l'EMD, chaque échantillon de la source (ligne) est apparié avec un seul échantillon de la cible (colonne) et dont le coefficient représentant son poids est maximal. Dans le cas de Sinkhorn, un échantillon de la source peut être apparié avec plusieurs échantillons cibles, qui ont des poids plus au moins élevés en fonction de leur influence lors du transport de l'échantillon source concerné. Les poids associés aux couplages optimaux sont représentés entre chaque paire d'échantillons avec des carrés de couleur selon la palette indiquée.

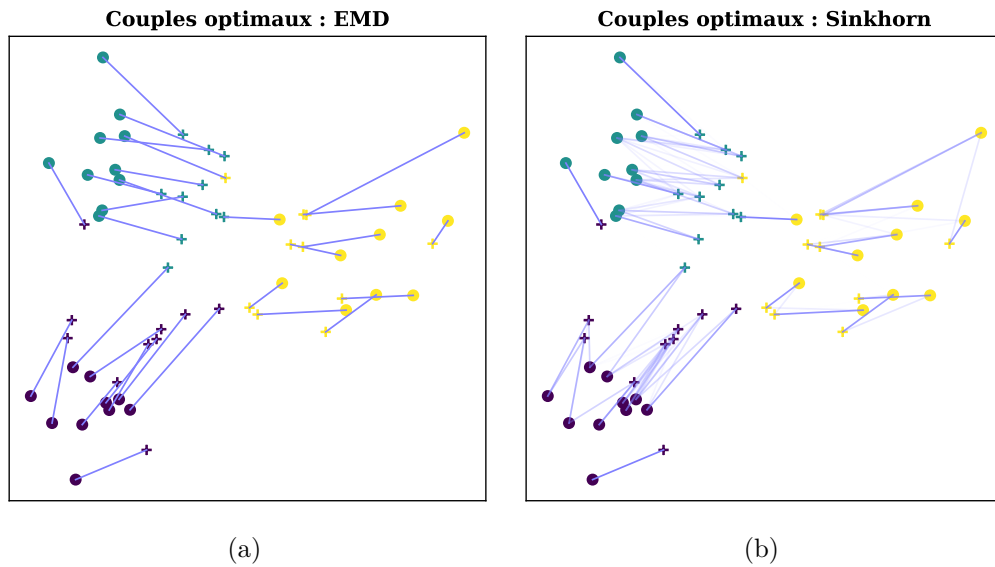


FIGURE 4.5 – Couplage entre les éléments de la source et de la cible avec l’EMD **(a)** et avec régularisation par l’entropie Sinkhorn **(b)**. Les liens en bleu montrent les couples optimaux formés où chaque échantillon de la source sera transporté vers le ou les échantillons de la cible associés. Ces liens sont plus denses avec la régularisation par l’entropie qu’avec l’approche EMD.

Le résultats des échantillons transportés est illustré dans la Figure 4.6. Étant donné qu’avec l’EMD, chaque échantillon de la source est couplé avec un seul échantillon de la cible et ayant un poids maximal (affichés en jaune dans la Figure 4.5a), tous les échantillons sources transportés coïncident parfaitement avec leurs homologues de la cible (Figure 4.6a). Avec Sinkhorn, l’effet de la pondération dû au couplage plus dense et traduisant l’influence de plusieurs échantillons cibles dans la manière dont l’échantillon source est transporté, se marque par un décalage entre les échantillons transportés et les échantillons de la cible (Figure 4.6b).

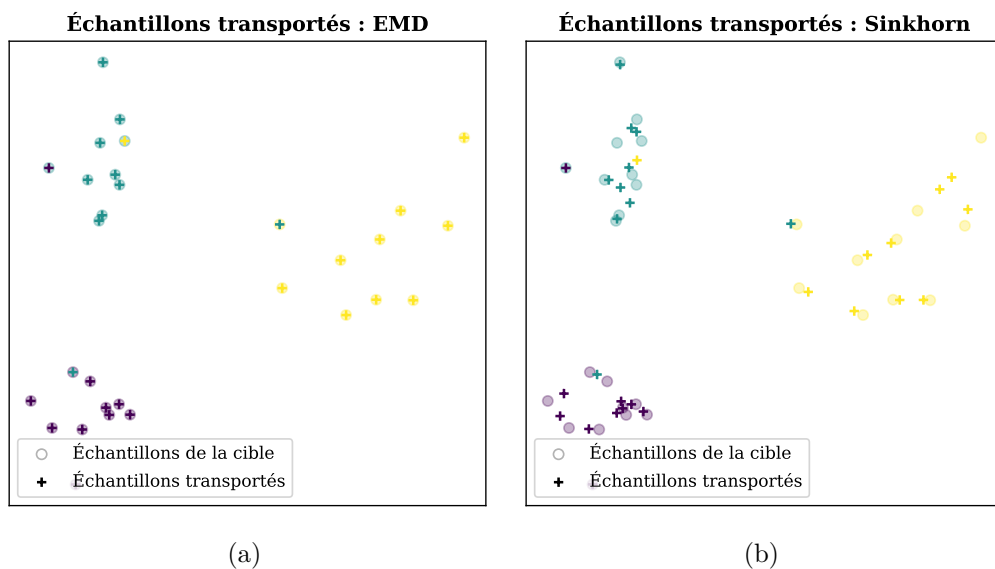


FIGURE 4.6 – Résultats des échantillons transportés avec **(a)** l’EMD et **(b)** Sinkhorn selon les couples optimaux formés dans la Figure 4.5.

Que ce soit avec l'algorithme de l'EMD ou Sinkhorn, seules les distributions de la source et de la cible ont été exploitées pour former les couples optimaux. Les informations sur les classes des échantillons n'ont jusqu'à présent servi qu'à des fins de visualisation.

- **Régularisation par les étiquettes :** Courty et al. (2014) ont proposé d'inclure un nouveau terme de régularisation dans l'équation de transport qui tient compte de l'information des étiquettes disponibles sur la source afin d'estimer une meilleure fonction de transport. Comme les étiquettes de la source sont généralement disponibles, il s'agit de considérer cette information afin de pénaliser les couplages qui associent des échantillons sources ayant des étiquettes différentes avec les mêmes échantillons de la cible (étiquettes inconnues), puisqu'un échantillon ne peut avoir qu'une seule étiquette. Le problème d'optimisation devient ainsi :

$$\gamma_0^\lambda = \operatorname{argmin}_{\gamma \in \mathcal{B}} \langle \gamma, \mathbf{C} \rangle_F + \lambda \Omega_s(\gamma) + \eta \Omega_c(\gamma) \quad (4.15)$$

où $\eta \geq 0$ est un paramètre qui pondère l'impact de la régularisation et $\Omega_c(\cdot)$ est un algorithme de régularisation basé sur les normes $\ell_2 - \ell_1$ ou $\ell_p - \ell_1$ (Courty et al., 2014).

Comme seules les étiquettes de la source sont impliquées, il s'agit d'une adaptation de domaine non supervisée. Une régularisation impliquant des étiquettes cibles a également été proposée. L'adaptation de domaine est dite semi-supervisée en exploitant un nombre réduit de ces étiquettes dans la cible. Ceci implique la proposition d'un nouveau terme de régularisation. Il est conçu de telle sorte que les échantillons du domaine cible ne doivent être appariés qu'avec des échantillons de la source ayant les mêmes étiquettes. La régularisation consiste à augmenter le coût des couplages qui ne vérifient pas cette contrainte. Ce terme de régularisation est exprimé comme suit :

$$\Omega_{semi}(\gamma) = \langle \gamma M \rangle \quad (4.16)$$

où M est une matrice de coût de taille $N_s \times N_c$, avec $M(i, j) = 0$ si $y_i^s = y_j^c$ ou si j est un échantillon de la cible sans étiquette connue, sinon $M(i, j) = +\infty$. Cela revient à modifier la fonction de coût initiale, définie dans l'équation 4.12 en affectant un coût infini aux couplages ne respectant pas cette contrainte.

Dans cette première partie théorique, nous avons présenté les concepts de base du transport optimal de manière générale et plus particulièrement, sa formulation pour l'adaptation de domaine. Dans ce deuxième cas, l'objectif est de modifier la distribution des données de la source de façon à ce qu'elle ressemble à celle de la cible afin d'entraîner ensuite un seul classifieur sur la source transportée qui serait plus performant sur la cible. Pour y répondre, quatre variantes du transport optimal sont envisagées :

- **EMD** : c'est une approche non supervisée dans laquelle seules les distributions de la source et de la cible sont considérées. On cherche à former des couples optimaux, qui minimisent le coût global du transport, entre chaque échantillon de la source et un seul échantillon de la cible. Une fois les couples formés, les échantillons sources sont déplacés vers leurs homologues de la cible.
- **Sinkhorn** : il s'agit toujours d'une approche non supervisée ne prenant en compte que les deux distributions source et cible. La seule différence avec l'approche précédente est la régularisation par l'entropie qui permet à un échantillon de la source d'avoir plus d'un homologue dans la cible. Les échantillons cibles affectent la fonction de transport et influencent la façon dont l'échantillon source est transporté. La nouvelle position de l'échantillon source ne coïncidera pas avec un seul échantillon de la cible mais sera fonction de tous les autres.
- **Sinkhorn avec les étiquettes de la source** : toujours en mode non supervisé, cette troisième variante consiste à incorporer les étiquettes des échantillons de la source pour une meilleure estimation de la fonction de transport. Partant du principe qu'un échantillon de la cible ne peut avoir qu'une seule étiquette, elle permet de pénaliser les couples qui associent des échantillons de la source avec des étiquettes différentes aux mêmes échantillons de la cible (étiquettes inconnues) et ce, en augmentant leur coût de transport.
- **Sinkhorn avec les étiquettes de la source et de la cible** : cette dernière variante est une approche semi-supervisée. En plus de considérer les étiquettes de la source, elle autorise l'inclusion de quelques étiquettes disponibles dans la cible. Cette fois-ci, elle pénalise les couples reliant des échantillons cibles avec des étiquettes connues à des échantillons sources qui n'ont pas les mêmes étiquettes.

4.2 Application du transport optimal pour l'identification des peupleraies sur de grandes étendues

En télédétection, le transport optimal peut être envisagé comme une méthode d'adaptation de domaine par adaptation des distributions. En effet, en considérant une série temporelle d'images satellitaires avec b bandes spectrales acquises à d dates, l'espace des caractéristiques sera formé pour tous les pixels disponibles avec $b \times d$ variables issues des bandes spectrales et temporelles formant l'image, ainsi que leur étiquettes relatives aux différentes classes qu'ils représentent. Il ne s'agit donc pas de déplacer des objets matériels d'une source vers une cible, mais de modifier les valeurs des pixels de la source pour qu'elles ressemblent à celles de la cible.

Pour la détection des peupleraies sur de grandes étendues à partir d'images satellitaires, on peut donc considérer deux domaines source et cible définis par deux séries temporelles identiques (même nombre de variables spectrales et temporelles) mais spatialement distinctes (*transfert spatial*). Dans ce mode de transfert, l'application du transport optimal permettrait de réduire le décalage entre les deux séries temporelles dû, notamment, à la variabilité spatiale de la réponse spectrale des peupleraies ou au décalage des dates clés du cycle phénologique. Il serait également possible d'appliquer le transport optimal en considérant deux séries temporelles sur la même zone géographique mais acquises à deux années différentes (*transfert temporel*). Bien que le principal objectif soit d'identifier les peupleraies sur de grandes étendues et qu'une approche de transfert spatial soit à envisager, cette question du transfert temporel s'est également présentée. Compte tenu de la difficulté de collecter des données de référence pour chaque année, nous nous sommes interrogés sur la possibilité d'envisager un modèle unique sur une année pour prédire différentes années de séries temporelles utilisant le transport optimal afin de compenser la variabilité temporelle de la réponse spectrale d'une année à l'autre. De même, un *transfert spatio-temporel* peut être envisagé. Le but étant d'explorer le potentiel du transport optimal face à une variabilité entre deux séries temporelles spatialement et temporellement distinctes.

À notre connaissance, seuls les travaux de (Tardy et al., 2019) ont abordé la question de l'adaptation de domaine par transport optimal avec des séries temporelles d'images satellitaires. Ils ont exploité sept séries temporelles d'images Formosat-2 pour effectuer un transfert temporel entre années, où les domaines source et cible désignaient à chaque fois deux années différentes.

Dans le cas de ces travaux, il s'agit d'explorer l'apport du transport optimal selon les trois modes de transfert :

- **transfert spatial** : dans lequel les domaines source et cible correspondent à deux séries temporelles Sentinel-2 de la même année mais issues de deux tuiles différentes,
- **transfert temporel** : où on considère une même tuile Sentinel-2 sur deux années différentes,
- **transfert spatio-temporel** : les deux domaines correspondent à deux séries temporelles issues de deux tuiles différentes et acquises à deux années différentes.

4.2.1 Protocole expérimental

Afin d'évaluer l'apport du transport optimal selon les trois modes prédéfinis, nous avons tout d'abord fixé arbitrairement un seul sens de transfert : depuis la tuile du Sud-Ouest (31TCJ) vers la tuile du Nord-Est (31UEQ). À chaque fois, il s'agit de transporter les données de la source (tuile 31TCJ) vers la cible (tuile 31UEQ). En d'autres termes, ce sont les valeurs

des pixels de la source qui seront adaptés. Pour rappel, les espaces de caractéristiques de la source et de la cible formés par les valeurs des pixels dans les différentes bandes spectrales ainsi que les étiquettes de leurs classes d'appartenance sont issus des pré-traitements détaillés dans la Section 2.2.3. Ensuite, un modèle de classification supervisée est entraîné avec les données de la source transportées et leurs étiquettes qui sont conservées après le transport en respectant la distribution conditionnelle (Équation (4.10)). Enfin, une fois le modèle appris sur les données transportées (donc comparables au domaine cible), il s'agit de prédire les étiquettes dans le domaine cible.

Rappelons que l'intuition derrière le transport optimal pour l'adaptation de domaine est que si les données sources subissent une transformation pour ressembler aux données cibles, le *datashift* est réduit, et un classifieur entraîné sur les données sources adaptées devrait être plus performant qu'un classifieur reposant uniquement sur les données originelles du domaine source.

Scénarios de transport

Avant de procéder au transport des distributions, il a fallu définir les classes à prendre en compte dans les deux domaines ainsi que leurs proportions. Comme l'hypothèse derrière le transport optimal est que chaque échantillon dans la source possède un homologue dans la cible, les classes doivent être les mêmes entre les deux domaines afin que le transport soit cohérent. Intuitivement, la première approche était de ne garder que les classes communes entre les deux domaines (scénario 1). Comme mentionné dans le Tableau 2.1, ces classes peuvent être des classes d'essences pures ou des classes mixtes. Cependant, comme les classes mixtes contiennent des mélanges de ces mêmes essences pures, cela peut impacter la qualité du transport en créant des couplages entre pixels purs et pixels issus de classes de mélange. Par conséquent, nous avons considéré un deuxième scénario (scénario 2) où seules les classes pures et communes aux deux domaines sont conservées. Enfin, un troisième scénario (scénario 3) a consisté à ne considérer que deux classes, (1) peuplier et (2) non-peuplier, la seconde contenant l'ensemble des autres classes d'essences pures. L'intuition derrière ce choix est que, comme c'est la classe des peupliers qui nous intéresse principalement, les erreurs de transport qui pourraient apparaître entre les autres espèces sont de moindre importance. Ces trois scénarios ont été testés dans le cas d'un transfert spatial, et seul le meilleur scénario a été adopté pour les deux autres modes de transfert temporel et spatio-temporel.

En ce qui concerne le nombre d'échantillons, Courty et al. (2014) recommandent que pour garantir les meilleurs résultats, les proportions d'échantillons de chaque classe doivent être similaires dans les distributions source et cible, avec un écart maximal de $\pm 15\%$ entre les deux domaines. Le problème ne s'est pas posé car les échantillons sont déjà sélectionnés de manière

équilibrée avec le même nombre par classe dans la même tuile (Section 2.4, Chapitre 2). ce qui permet de garder des proportions similaires des classes entre deux tuiles. Afin de pouvoir comparer les performances des trois scénarios et de maintenir l'équilibre des proportions entre les classes, nous avons considéré tous les échantillons de peuplier disponibles dans les ensembles d'entraînement de la source avec 3850 pixels et de la cible avec 1250 pixels. Ce nombre d'échantillons est ensuite réparti entre les autres classes. Pour le scénario 1, cela représente 962 et 312 échantillons pour chacune des quatre autres classes de la source et de la cible respectivement. Dans le scénario 2, le nombre d'échantillons est de 1925 et 625 pour chacune des deux classes considérées dans la source et dans la cible. Enfin dans le scénario 3, la classe non-peuplier est composée de 1283 échantillons pour la source et 417 pour la cible pour chacune des trois classes disponibles. Chaque jeu d'échantillon est ainsi composé de 50% d'échantillons de peuplier et de 50% d'échantillons répartis entre les autres classes définies par chaque scénario (Tableau 4.2).

Modes de transport

Le transport optimal est appliqué entre deux séries temporelles de tuiles Sentinel-2 selon les trois modes prédéfinis. Les séries temporelles de 2017 sont constituées de 340 bandes relatives aux 34 dates multipliées par les 10 bandes spectrales par date. Bien que les séries temporelles de 2018 soient composées de 36 dates, seules les 34 dates communes avec l'année 2017 ont été conservées afin de préserver les mêmes dimensions dans les cas des transferts temporel et spatio-temporel.

Le transfert spatial est effectué depuis la tuile source du Sud-Ouest (31TCJ) de 2017 vers la tuile cible du Grand-Est (31UEQ) de la même année. Pour le transfert temporel, nous avons considéré la tuile du Sud-Ouest avec la série temporelle de 2017 comme source et celle de 2018 comme cible. Dans le cas du transfert spatio-temporel, la tuile de Sud-Ouest de 2017 demeure, comme dans les deux modes précédents, la source et la série temporelle de la tuile du Grand-Est de 2018 est considérée comme la cible. Dans les trois modes, toutes les séries temporelles comptent 340 bandes spectrales.

Transport optimal non-supervisé et semi-supervisé

Pour chacun des trois modes, le transport optimal est appliqué selon ses quatre variantes non-supervisée ou semi-supervisée en amont de l'étape classique de classification supervisée. Les deux espaces de caractéristiques de la source (Ω_s) et de la cible (Ω_c) utilisés en entrée du transport optimal sont issus des pré-traitements des séries temporelles détaillés dans la Section 2.2.3. Ces étapes sont récapitulées dans la Figure 4.7.

En mode non-supervisé, nous avons testé l'approche avec l'**EMD** en ne considérant que

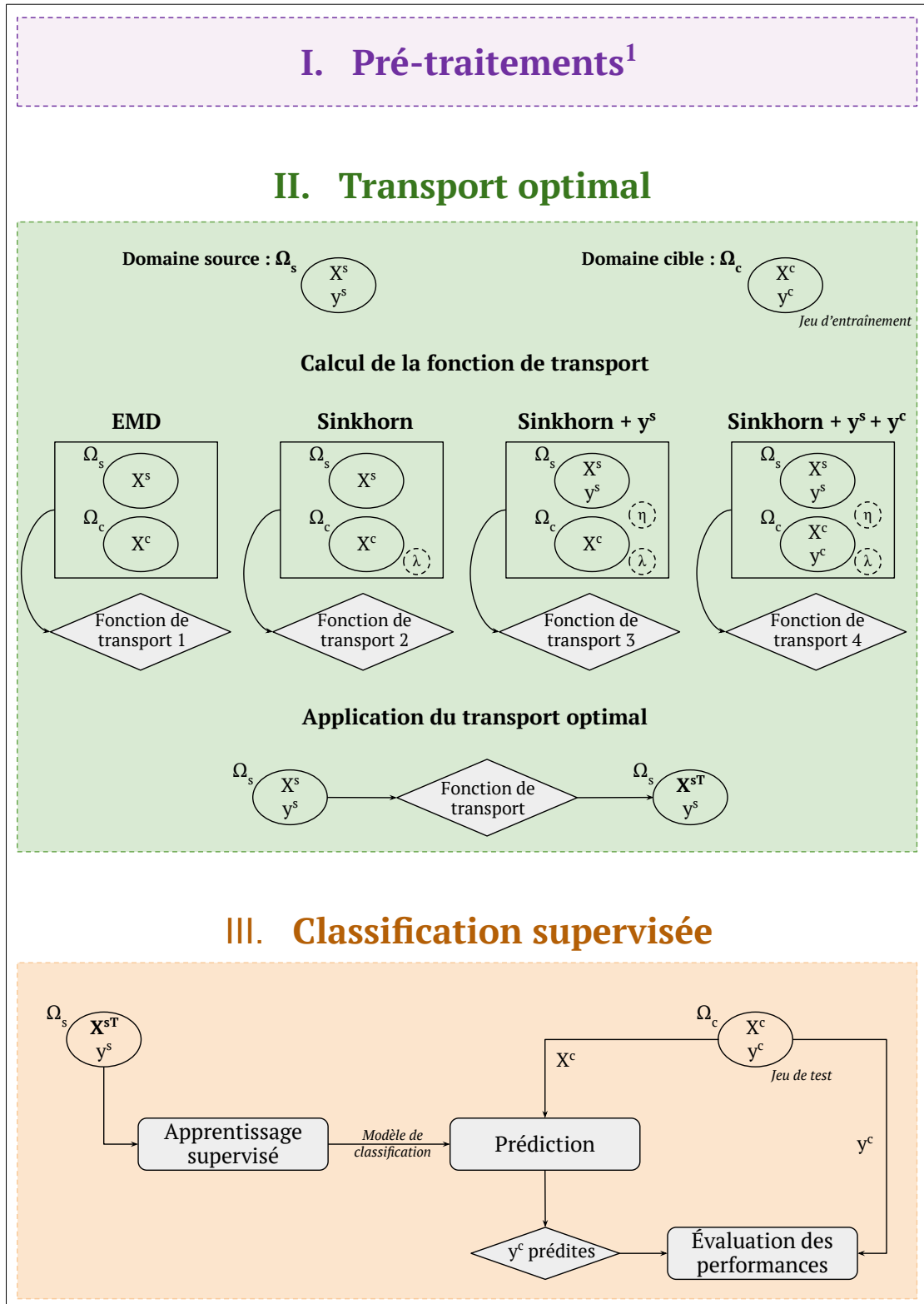


FIGURE 4.7 – Schéma des quatre variantes du transport optimal envisagées en amont de la classification de la cible. Les domaines source et cible sont issus des pré-traitements (1) détaillés dans la Section 2.2.3. Les deux paramètres λ et η désignent respectivement les deux termes de régularisation par l'entropie et par les étiquettes.

les deux distributions source et cible. Nous avons également appliqué l’approche régularisée par l’entropie avec l’algorithme de **Sinkhorn**. Enfin nous avons considéré la régularisation par les étiquettes des échantillons du domaine source **Sinkhorn + y^s** .

En mode semi-supervisé, nous avons testé l’approche **Sinkhorn + y^s + y^c** dans laquelle les étiquettes des échantillons de la source et celles de quelques échantillons de la cible sont prises en compte dans le transport optimal. Dans cette dernière approche semi-supervisée, nous avons évalué l’impact du nombre d’étiquettes cibles rajoutées. Nous avons ainsi testé différents pourcentages du nombre d’étiquettes parmi les échantillons d’entraînement disponibles : [1, 5, 10, 15, 25, 35, 50, 75, 85, 100].

Démarche de classification et de validation

Le transport optimal constitue une phase préliminaire à la classification supervisée. En effet, une fois la fonction de transport calculée et les échantillons sources transportés, le classifieur est entraîné à partir de ces échantillons sources transportés et de leurs étiquettes, puis les échantillons cibles sont prédits et la performance est évaluée.

Pour le domaine cible, deux jeux d’échantillons ont été considérés. Un premier jeu d’« entraînement » qui n’est pas impliqué dans la classification supervisée mais qui est utilisé dans le calcul de la fonction de transport et lors de la régularisation, et un jeu de test qui intervient seulement dans l’étape de l’évaluation de la classification.

Dans les trois variantes régularisées du transport optimal (Sinkhorn, Sinkhorn + y^s et Sinkhorn + y^s + y^c), les paramètres de régularisation du transport (λ et η) ont été fixés selon une grille de recherche par validation croisée *5 folds* (Section 3.2.2). Un total de 20 valeurs comprises entre 0.001 et 1000 ont été évaluées pour λ et η . Après transport, un classifieur RF est entraîné sur les données sources transportées dont les hyper-paramètres ont également été optimisés par validation croisée *5 folds* (voir la Section 3.2.2 pour la liste des hyper-paramètres et des valeurs testées). Les performances de classification sont évaluées sur le jeu de test indépendant de la cible. Les valeurs d’OA et de PA sont respectivement utilisées comme métriques de performance globale et par classe.

Afin d’évaluer la contribution de l’adaptation de domaine par transport optimal, les performances sont à chaque fois comparées à celles d’un classifieur entraîné sur les mêmes données sources non transportées.

L’ensemble ces traitements ont été implémentés en Python avec la version 0.6 de la bibliothèque *Python Optimal Transport* (POT) (Flamary et al., 2021), qui regroupe tous les algorithmes de transport régularisé et non régularisé, ainsi que *scikit-learn* (Pedregosa et al., 2011) pour la classification et l’optimisation des paramètres.

4.2.2 Transfert spatial

Pour un transfert spatial depuis la tuile source du Sud-Ouest (31TCJ) vers la tuile cible du Nord-Est (31UEQ) nous avons considéré les séries temporelles de 2017. Ainsi, le nombre d'échantillons d'entraînement disponible est de 3850 et 1250 échantillons par classe pour la source et la cible où chaque échantillon est caractérisé par une classe de feuillus et 340 variables spectrales et temporelles (10 bandes spectrales \times 34 dates). Le nombre de classes et le nombre d'échantillons sélectionnés selon chaque scénario sont récapitulés dans le Tableau 4.2.

TABLEAU 4.2 – Nombre d'échantillons par classe selon le scénario considéré dans le transfert spatial. Dans le scénario 1, toutes les classes pures et mixtes communes aux deux domaines sont conservées. Dans le scénario 2, seules les classes pures et communes sont retenues. Dans le scénario 3, deux classes sont considérées avec une classe spécifique aux peupliers et une classe contenant toutes les autres essences pures disponibles dans chaque domaine.

| Scénario | Domaine | Peuplier | Robinier | Châtaignier | Chêne | Hêtre | Forêt fermée | Forêt ouverte |
|------------------|---------|----------|-------------------------|-------------|----------|----------|--------------|---------------|
| Scénario 1 | Source | 3850 | 962 | - | 962 | - | 962 | 962 |
| | Cible | 1250 | 312 | - | 312 | - | 312 | 312 |
| 5 classes | | classe 1 | classe 2 | - | classe 3 | - | classe 4 | classe 5 |
| Scénario 2 | Source | 3850 | 1925 | - | 1925 | - | - | - |
| | Cible | 1250 | 625 | - | 625 | - | - | - |
| 3 classes | | classe 1 | classe 2 | - | classe 3 | - | - | - |
| Scénario 3 | Source | 3850 | 1283 | 1283 | 1283 | - | - | - |
| | Cible | 1250 | 417 | - | 417 | 417 | - | - |
| 2 classes | | classe 1 | classe 2 ⁽¹⁾ | classe 2 | classe 2 | classe 2 | - | - |

⁽¹⁾ La classe 2 du scénario 3 est formée par tous les échantillons de classes pures disponibles autres que le peuplier et qui peuvent être différentes entre la source et la cible.

Analyse des trois scénarios

Les résultats du transport optimal spatial dans les quatre configurations et selon les trois scénarios sont résumés dans le Tableau 4.3. L'application du transport optimal avec une approche non régularisée utilisant l'EMD a généralement conduit à une amélioration de la précision globale de la classification sur la cible de 3 à 6% et une augmentation du PA peuplier de 11 à 16% en comparaison à un modèle entraîné sur les données sources originelles non transportées (t_0). La valeur de PA obtenue pour la classe peuplier varie de 85 à 88% entre les trois scénarios. Cela n'a pas été le cas pour les autres classes où les valeurs de PA se sont parfois dégradées après le transport comme la classe des forêts fermées (-39%) dans le scénario 1 ou la classe non-peuplier (-9%) dans le scénario 3.

Pour les deux premiers scénarios, la régularisation par l'entropie avec Sinkhorn, qui se traduit par une augmentation de la densité des couplages entre la source et la cible, n'a

été que légèrement meilleure en termes d'OA et de PA peuplier par rapport à l'approche non régularisée. L'amélioration était plus nette pour le scénario 3 avec +7% pour le PA de la classe peuplier et +2% pour l'OA. Le PA peuplier était de 92% marquant une différence de 4 à 9% entre les trois scénarios. Comme nous avons les mêmes échantillons de peuplier dans les trois scénarios, cet écart est un peu surprenant, car on s'attendait à avoir des résultats équivalents comme pour le cas non régularisé (EMD). En ce qui concerne les autres classes, notamment le robinier et le chêne, les valeurs de PA se sont dégradées. En examinant les matrices de confusion issues du scénario 1, l'utilisation de Sinkhorn a entraîné une plus grande confusion entre les échantillons de robinier et de chêne d'une part et ceux des chênes et de la classe des forêts fermées d'autre part (Figure 4.8a et 4.8b). Bien qu'aucune information sur les étiquettes n'ait été introduite pour cette configuration (Sinkhorn), la nature des classes présentes avec le peuplier semble influencer l'estimation de la fonction de transport et ainsi, la qualité de la classification.

TABLEAU 4.3 – Résultats des quatre configurations de transport optimal selon les trois scénarios de classes selon un mode de transfert spatial depuis la tuile du Sud-Ouest (31TCJ) vers la tuile du Nord-Est (31UEQ). La colonne t_0 reporte les performances d'une classification supervisée à partir des données sources non transportées. Les performances sont données en termes de précision globale (OA) et de précision producteur par classe (PA).

| Scénario | Performance | t_0 | EMD | Sinkhorn | Sinkhorn + y^s | Sinkhorn + y^s + y^c ⁽¹⁾ |
|------------|------------------|-------|-----|----------|---------------------|--|
| Scénario 1 | OA | 62 | 65 | 65 | 66 | 87 |
| | PA peuplier | 75 | 86 | 83 | 87 | 94 |
| | PA robinier | 34 | 39 | 24 | 35 | 81 |
| | PA chêne | 14 | 36 | 27 | 33 | 78 |
| | PA forêt fermée | 61 | 22 | 46 | 34 | 74 |
| | PA forêt ouverte | 86 | 74 | 87 | 78 | 91 |
| Scénario 2 | OA | 63 | 69 | 70 | 74 | 92 |
| | PA peuplier | 72 | 88 | 88 | 88 | 92 |
| | PA robinier | 50 | 52 | 49 | 53 | 93 |
| | PA chêne | 55 | 50 | 55 | 68 | 91 |
| Scénario 3 | OA | 87 | 87 | 89 | 90 | 96 |
| | PA peuplier | 74 | 85 | 92 | 95 | 96 |
| | PA non-peuplier | 99 | 90 | 85 | 85 | 96 |

⁽¹⁾ 100% des étiquettes cibles disponibles.

L'ajout de l'information des étiquettes de la source (Sinkhorn + y^s) est avantageux pour les classes d'essences pures, notamment le peuplier, mais pas pour les classes plus hétérogènes (classes des forêts mixtes ou non-peuplier) où la valeur de PA a soit diminué, soit stagné. L'information des étiquettes de la source est censée réduire le nombre de permutations possibles en pénalisant les couplages incohérents vis-à-vis des classes associées. Ce résultat, comme le précédent, met en évidence une certaine similarité entre les classes, mesurée ici en termes de

distance Euclidienne, qui a perturbé l'appariement des échantillons entre la source et la cible. À l'inverse, l'incorporation des étiquettes de la source a réduit cette confusion qui s'est traduite par une augmentation des valeurs de PA des classes pures. La dégradation des valeurs de PA des deux classes de mélange considérées dans le scénario 1 s'explique par la diminution du nombre d'échantillons correctement classés (valeurs en rouge dans les Figures 4.8b et 4.8c) en raison de leur ressemblance avec les classes pures dans lesquelles ils ont été affectés.

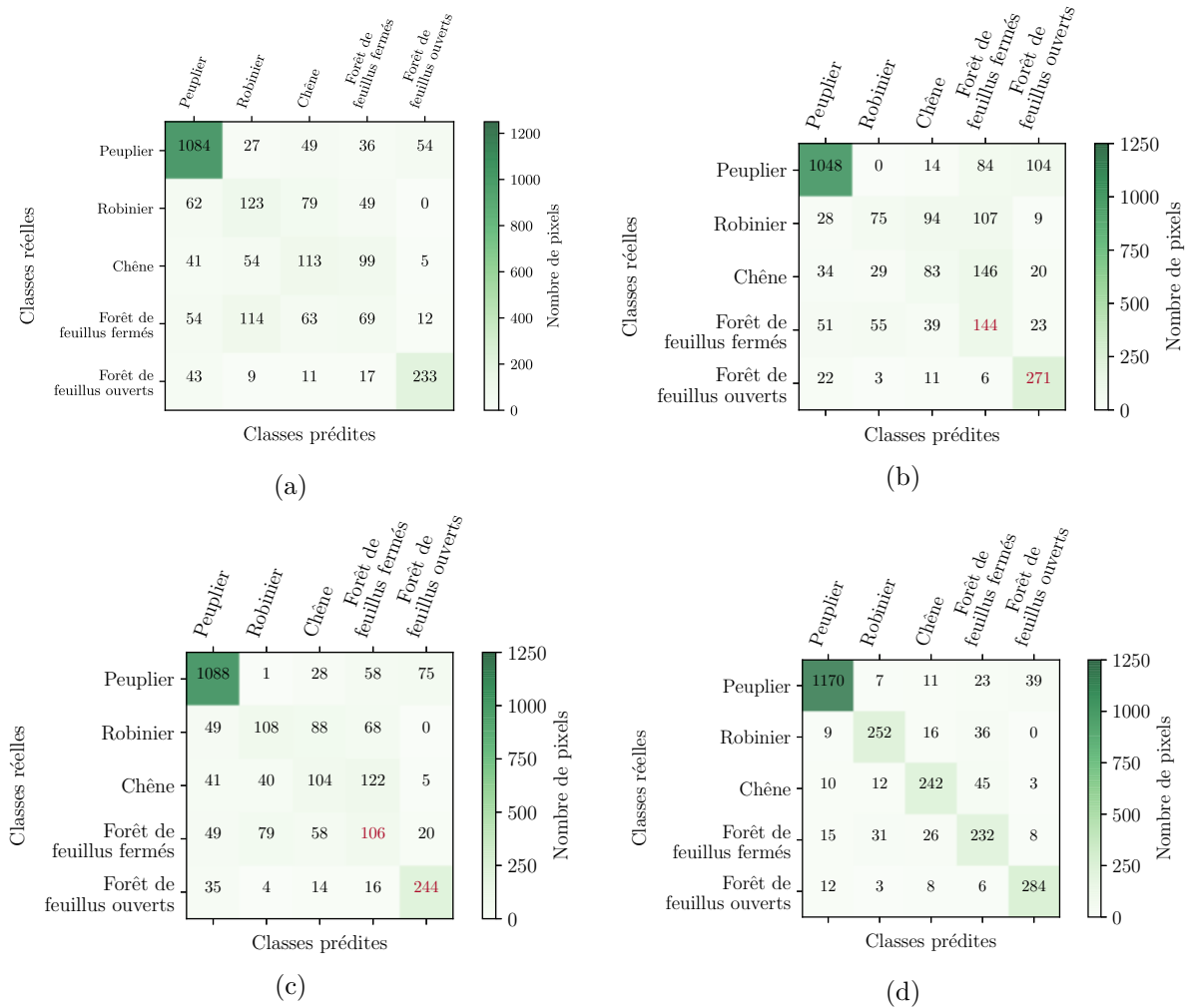
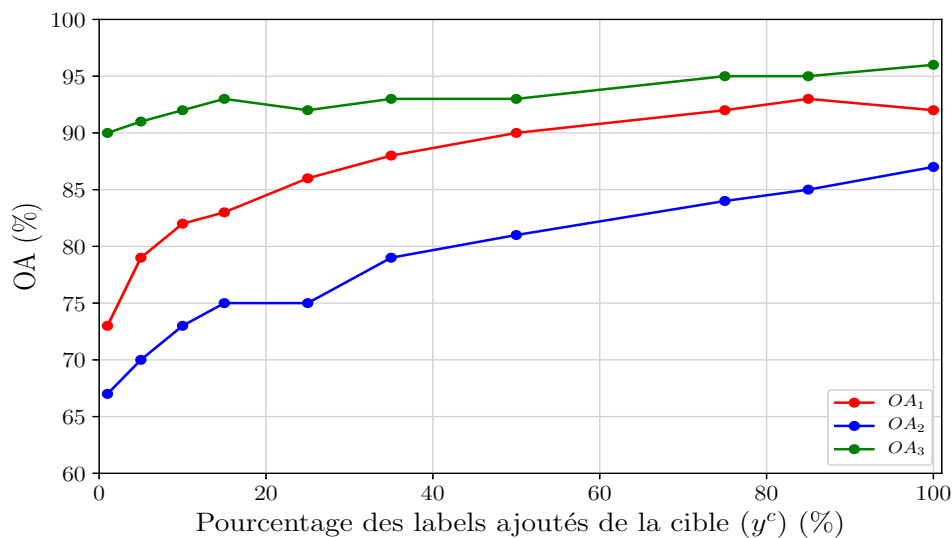


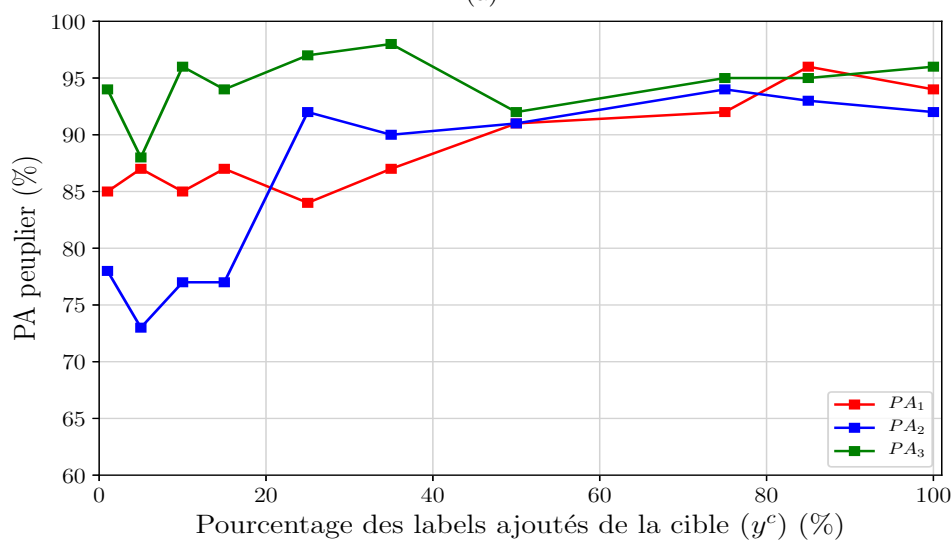
FIGURE 4.8 – Matrices de confusion calculées pour le scénario 1 de transfert spatial, avec la configuration (a) EMD, (b) Sinkhorn, (c) Sinkhorn + y^s et (d) Sinkhorn + y^s + y^c

Pour les trois scénarios, les meilleures performances sont atteintes avec la dernière configuration semi-supervisée incorporant des connaissances *a priori* sur les étiquettes des échantillons sources et de la cible (Sinkhorn + y^s + y^c). Les valeurs de PA obtenues pour les peupliers sont entre 92 et 96%. De même, les performances globales données par l'OA et les PA des autres classes ont atteint leurs valeurs maximales avec cette configuration qui implique l'ensemble des étiquettes cibles disponibles (100% des y^c du jeu d'entraînement).

L'analyse de l'impact du nombre d'échantillons étiquetés du domaine cible a montré qu'au-



(a)



(b)

FIGURE 4.9 – Évolution des performances en fonction du nombre d'étiquettes cibles y^c rajoutées dans la configuration du transport optimal semi-supervisé (Sinkhorn $+y^s + y^c$) pour chacun des trois scénarios. En (a) avec la précision globale (OA) et en (b) avec la PA de la classe peuplier.

tour de 50% de l'effectif (soit environ 625 pixels), la valeur de PA commençait à se stabiliser, quel que soit le scénario (Figure 4.9). La fluctuation observée au début de l'évolution peut s'expliquer par le fait que les étiquettes ont été sélectionnées de manière aléatoire. De plus, le tirage n'a pas été répété pour fournir des évaluations moyennes, ce qui rend l'interprétation un peu délicate. Par ailleurs, l'augmentation du nombre d'étiquettes a conduit à une meilleure convergence des trois scénarios. Bien que cette configuration soit celle qui garantit les meilleures performances, elle implique d'avoir un certain nombre d'échantillons étiquetés dans la cible ce qui n'était pas forcément le but initial.

L'amélioration des performances en fonction de l'ajout des termes de régularisation est plus rapide dans le scénario 3, à la fois de manière globale (OA) ou en termes de PA de la classe

peuplier. Par ailleurs, l'agrégation de différentes classes pures au sein d'une même classe non-peuplier ne semble pas altérer les performances de la classe peuplier. Bien que l'information des PA par classe soit riche et plus informative sur l'impact et le fonctionnement du transport optimal, le scénario 3 impliquant deux classes a fourni le meilleur résultat et semble plus pertinent pour la classe peuplier.

Comparaison des quatre variantes du transport optimal avec le scénario 3

Afin de mieux comprendre le fonctionnement du transport optimal et d'analyser les valeurs de performance atteintes pour la classe peuplier avec chacune des quatre variantes, nous avons tracé les profils temporels de la source et de la cible dans trois bandes spectrales issues des trois régions du visible (Bande 4, $\mu = 665 \text{ nm}$), du *Red edge* (Bande 5, $\mu = 705 \text{ nm}$) et du moyen infrarouge (Bande 11, $\mu = 1610 \text{ nm}$). Ils sont illustrés dans la Figure 4.10 avant l'application du transport optimal.

L'analyse de ces profils confirme tout d'abord la non-stationnarité spatiale de la réponse spectrale des peupleraies entre les deux tuiles source du Sud-Ouest et cible du Grand-Est. Pour les trois bandes, la différence est un peu plus marquée au début (janvier à mars) et à la fin (novembre et décembre) de l'année avec des valeurs de réflectance plus élevées dans la source. Durant ces deux périodes, les canopées sont ouvertes et leur couverture est minimale étant donné l'absence des feuilles. Par conséquent, les différences ne peuvent pas être liées aux variations du cycle phénologique des peupleraies entre les deux régions, mais principalement à des différences dans la végétation du sous-étage présente en dessous. En outre, quelques différences dans le comportement spectral sont observées pendant la saison de végétation, entre fin mars et début octobre.

L'application du transport optimal est censée adapter la signature spectrale de la source afin qu'elle se rapproche de celle de la cible. Un exemple de profils temporels obtenus après application du transport optimal avec l'approche EMD est présenté dans la Figure 4.11. Dans les trois bandes considérées, le comportement spectral des données sources transportées coïncide parfaitement avec celui de la cible. Cela signifie qu'une grande majorité des échantillons de peuplier de la source ont été appariés avec des échantillons de peuplier de la cible, puis, au moment du transport, leurs valeurs spectrales ont été modifiées pour correspondre à celles de leurs homologues de la cible affichant ainsi une signature spectrale presque identique. L'examen des couplages optimaux formés avec l'approche EMD confirme ce constat (Figure 4.12a). En effet, la plupart des échantillons de peuplier de la source (lignes) ont des coefficients élevés avec les échantillons de peuplier de la cible (colonnes) représentés en vert foncé traduisant de faibles coûts de transport entre ces échantillons. Cependant, la couleur vert clair indique des coefficients faibles reflétant des coûts de transport plus élevés entre la majorité des peupliers

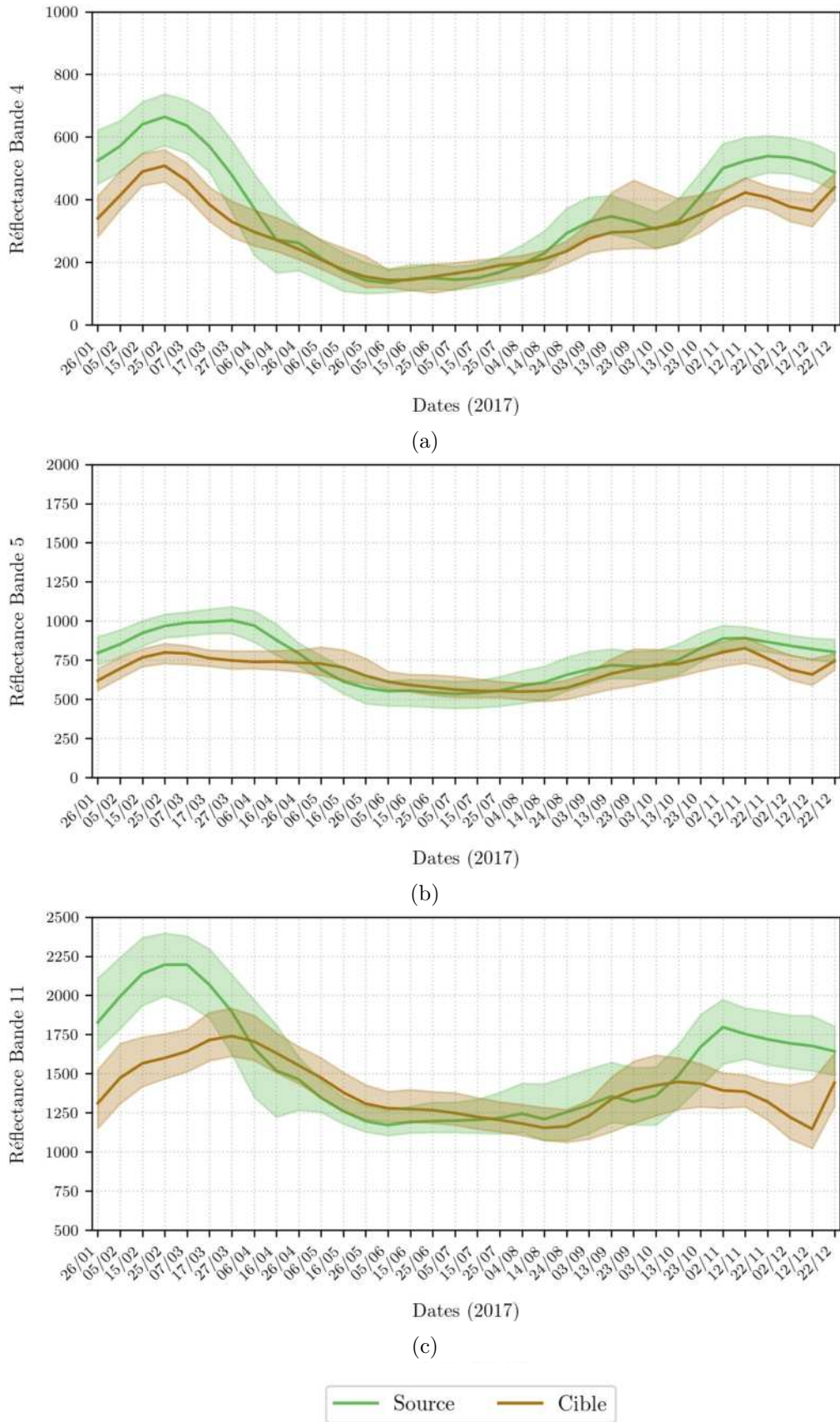


FIGURE 4.10 – Profil temporel de la classe peuplier dans la source (tuile 31TCJ) et dans la cible (tuile 31UEQ) avant transport optimal avec les séries temporelles de 2017 des (a) Bande 4, (b) Bande 5 et (c) Bande 11.

de la source et les échantillons cibles de la classe non-peuplier. Bien que la présence de certains couples optimaux entre les peupliers sources et les non-peupliers cibles n'ait pas affecté la signature médiane des données sources transportées dans les trois bandes considérées dans la Figure 4.11, elle peut expliquer la valeur de PA de la classe peuplier à 85% qui représente tout de même une amélioration de +11% par rapport à la situation sans transport optimal (colonne t_0).

L'augmentation du nombre de couples optimaux avec l'approche Sinkhorn associant des coefficients très élevés entre les échantillons sources et cibles de la même classe (Figure 4.12b) s'est traduite par une augmentation de +7% de la valeur de PA de la classe peuplier. La diminution de celle de la classe non-peuplier par rapport à l'approche EMD est due à l'augmentation du nombre de couples optimaux entre les échantillons de non-peuplier de la source et des échantillons de peuplier de la cible. Il en résulte une augmentation de la sur-détection des non-peupliers en peupliers, comme indiqué dans la matrice de confusion de la Figure 4.13b.

La pénalisation des couples associant des échantillons sources d'étiquettes différentes aux mêmes échantillons cibles avec l'ajout des étiquettes de la source (Sinkhorn + y^s), a permis une amélioration du PA peuplier de 3% mais elle n'a eu aucun effet sur celle de la classe non-peuplier. Les couples pénalisés se voient attribuer des coefficients très faibles (associés à des coûts élevés) indiqués en vert clair en Figure 4.12c.

La dernière variante incorporant les étiquettes sources et cibles (100% des étiquettes cibles y^c du jeu d'entraînement) dans le transport optimal (Sinkhorn + y^s + y^c) a abouti à des couplages optimaux entre échantillons source et cible appartenant aux mêmes classes tout en attribuant des coefficients nuls entre les échantillons de classes différentes 4.12d. Cela explique ainsi l'augmentation des performances pour les deux classes.

Nous avons également comparé les résultats cartographiques obtenus avec les quatre modèles issus des quatre variantes du transport optimal et utilisés pour prédire la tuile cible. Un extrait sur le département de l'Aisne est illustré dans la Figure 4.14.

Nous avons observé des discordances entre les résultats statistiques et les rendus cartographiques. Tout d'abord, avant l'application du transport optimal, la cartographie (Figure 4.14b) est très conforme avec la photo-interprétation de l'orthophotographie (Figure 4.14a). Les peupleraies sont globalement très bien identifiées avec de très faibles omissions et commissions. Malgré l'amélioration des performances statistiques de la classe peuplier avec les approches EMD, Sinkhorn et Sinkhorn + y^s , le résultat cartographique semble moins cohérent avec beaucoup de sur-détections de peupleraies. Ces commissions se révèlent plus limitées avec la dernière approche Sinkhorn + y^s + y^c .

Le dégradation du résultat cartographique illustrée par l'augmentation de la sur-détection des peupleraies semble être liée à la diminution du PA de la classe non-peuplier. Plus cette

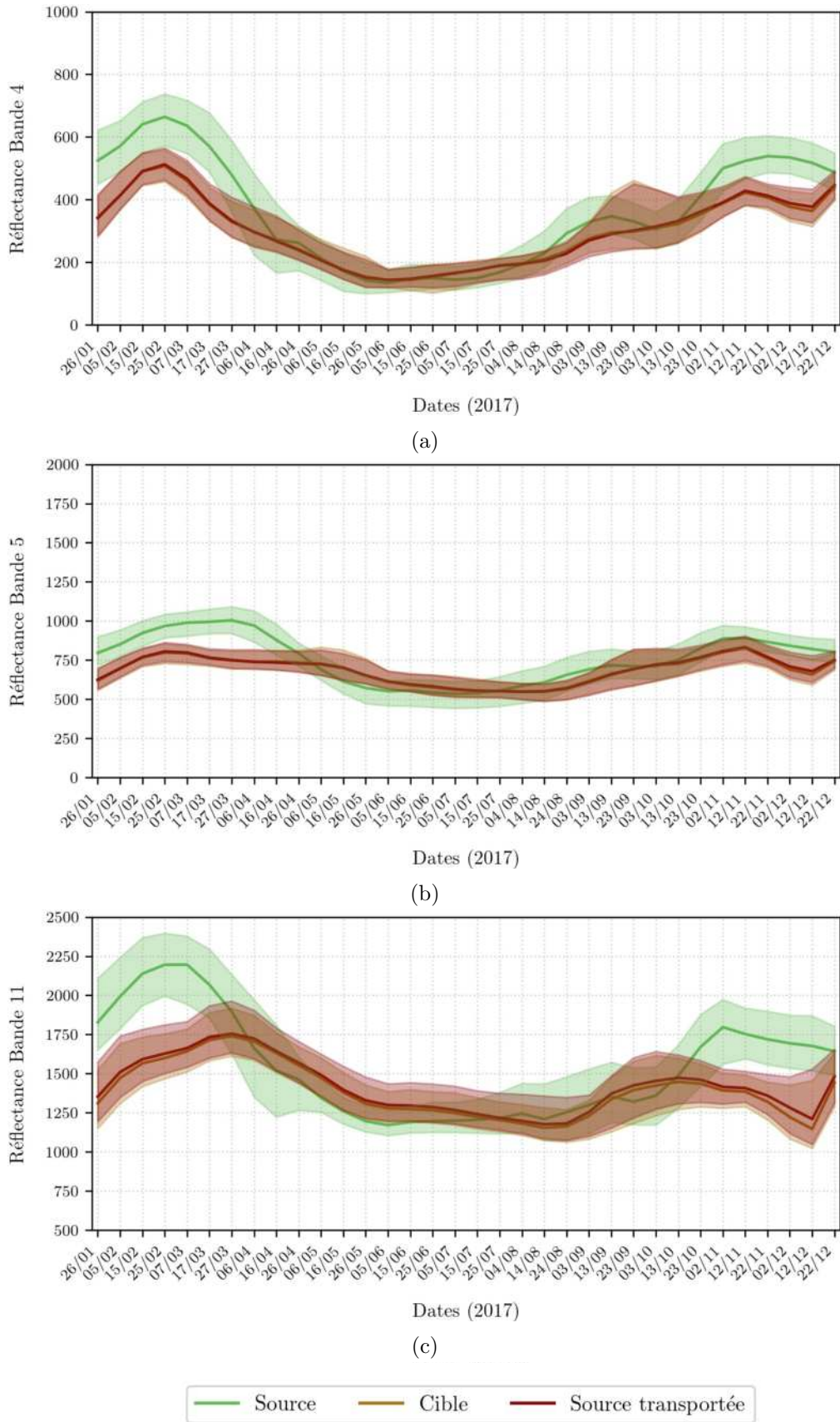


FIGURE 4.11 – Profil temporel de la classe peuplier dans la source (tuile 31TCJ), dans la cible (tuile 31UEQ) et dans la source transportée avec l’EMD avec les séries temporelles de 2017 de (a) Bande 4, (b) Bande 5 et (c) Bande 11.

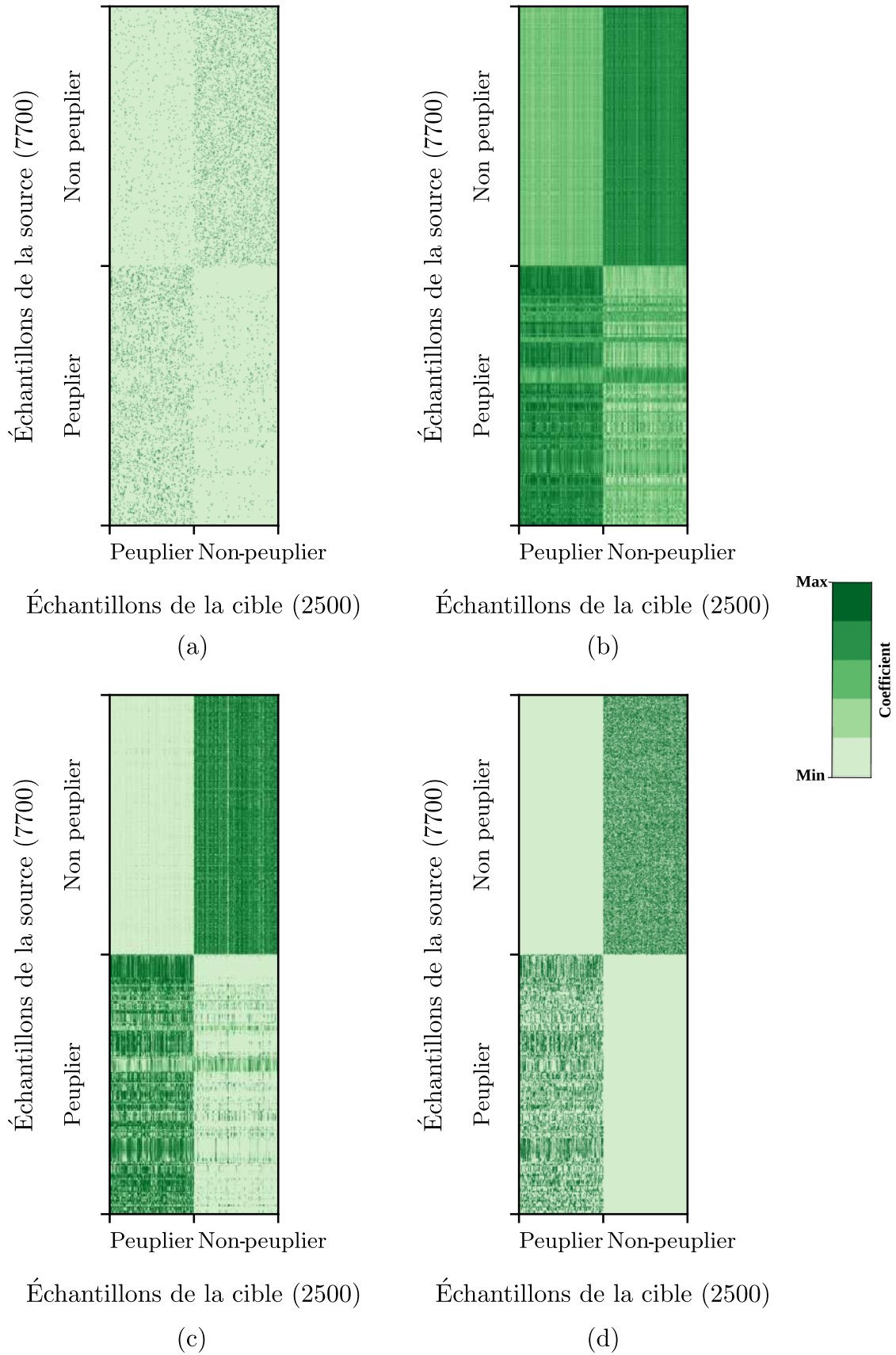


FIGURE 4.12 – Matrices des couplages optimaux entre les échantillons de la sources (lignes) et les échantillons de la cible (colonnes) dans le cas du transfert spatial lors le scénario 3 avec (a) EMD, (b) Sinkhorn, (c) Sinkhorn + y^s et (d) Sinkhorn + $y^s + y^c$.

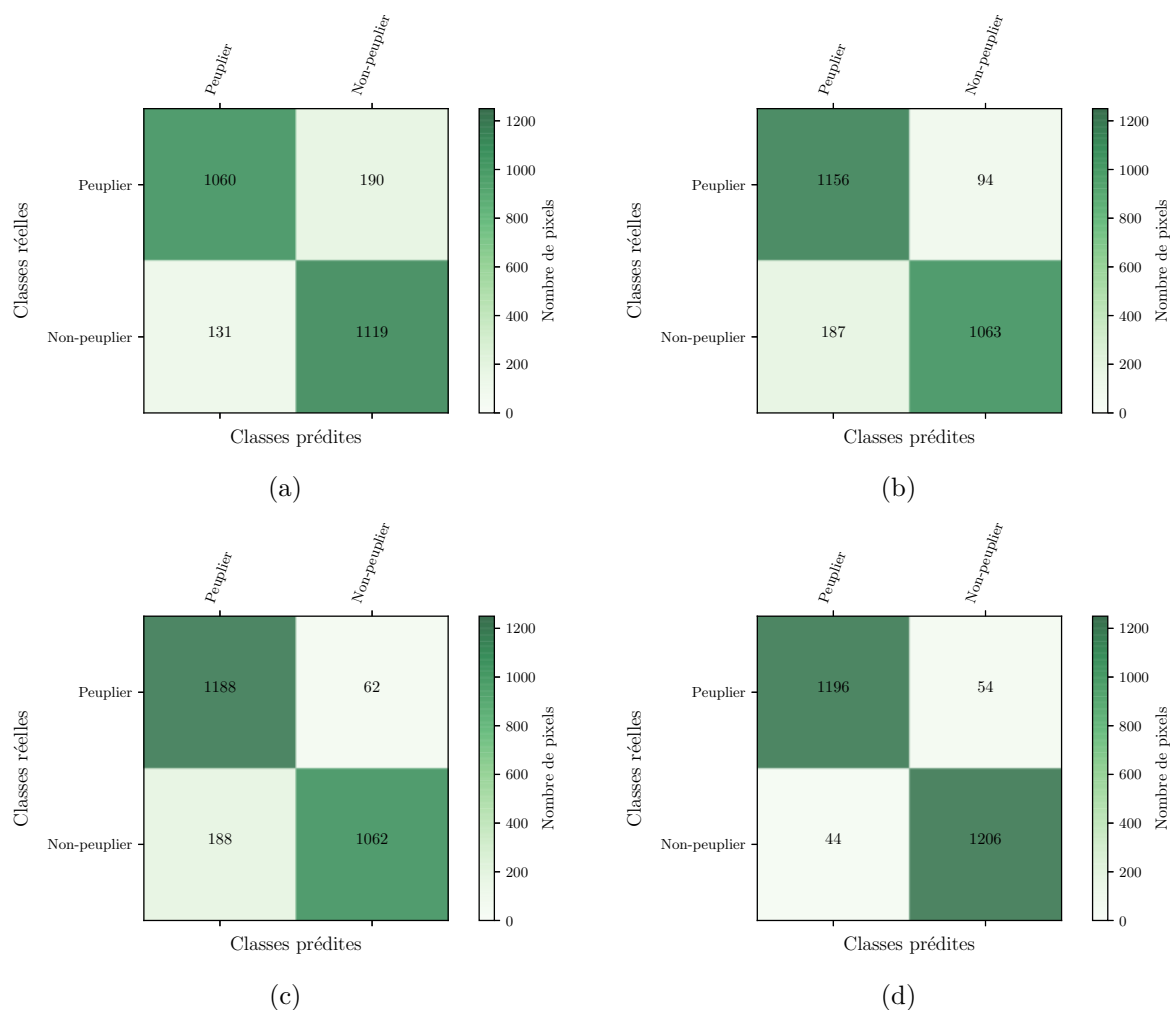


FIGURE 4.13 – Matrices de confusion calculées pour le scénario 3 de transfert spatial, avec la configuration (a) EMD, (b) Sinkhorn, (c) Sinkhorn + y^s et (d) Sinkhorn + $y^s + y^c$

valeur est faible, plus la sur-détection est importante. Cela interroge sur le choix de la classe non-peuplier, qui est très hétérogène et qui regroupe diverses classes de feuillus qui ne sont pas nécessairement communes à la source et à la cible. Nous avons examiné certaines commissions et une partie d'entre elles correspondait à des peuplements de hêtre (un exemple est donnée par la parcelle délimitée en jaune dans la Figure 4.14). Comme il s'agit d'une classe présente uniquement dans la cible, la confusion peut être justifiée par une plus grande similitude du comportement spectral avec les peupliers qu'avec les autres feuillus de la classe non-peuplier.

Dans cette partie, nous avons montré le potentiel de l'adaptation de domaine par transport optimal pour l'amélioration de la reconnaissance des peupleraies dans le cas d'un transfert spatial entre deux séries temporelles de tuiles Sentinel-2 spatialement distinctes et acquises la même année (2017). Nous avons d'abord comparé trois scénarios pour le choix des classes à prendre en compte dans le transport optimal. Nous avons déduit que l'augmentation du nombre de classes dans le jeu de données et la présence de classes de mélange freinaient

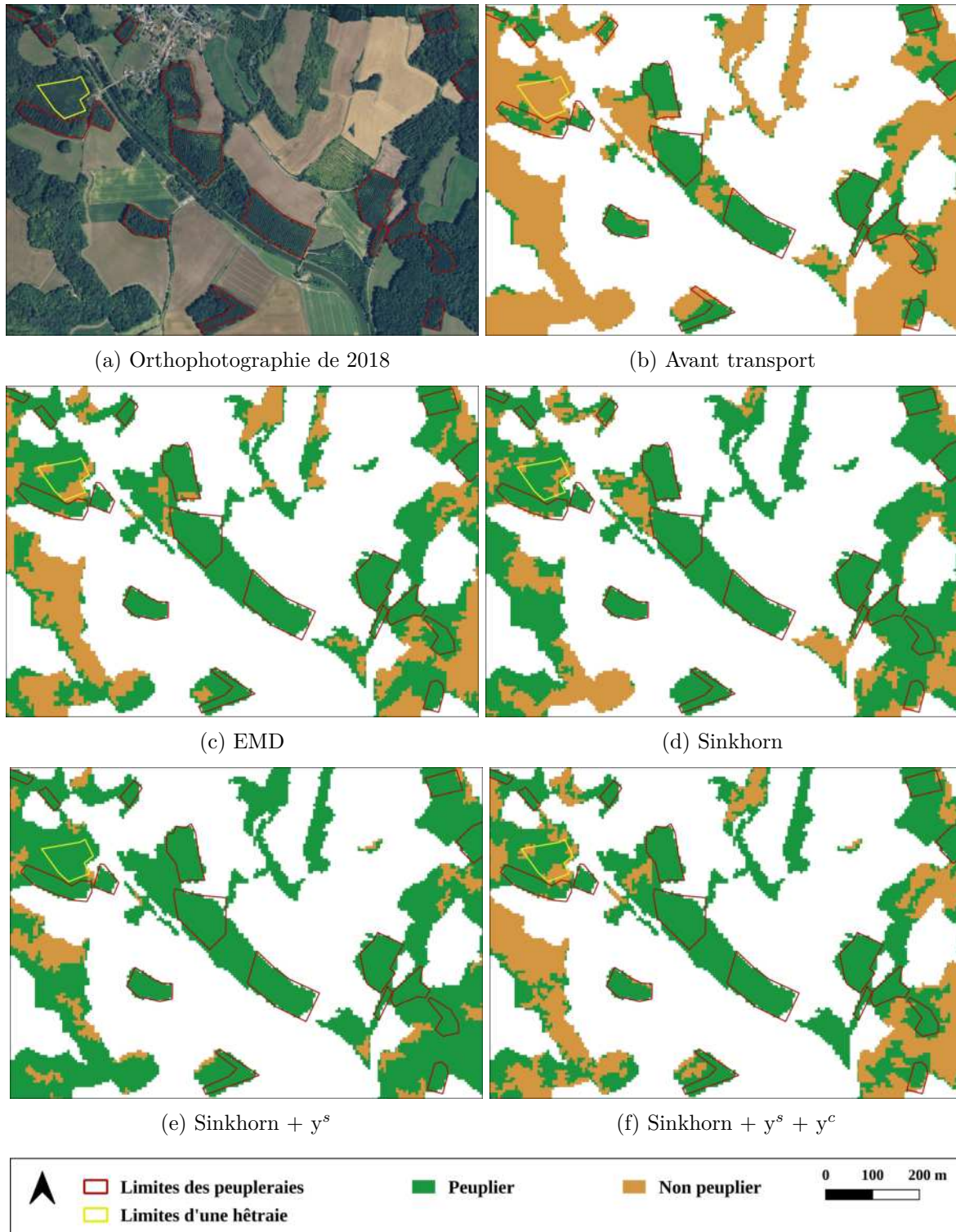


FIGURE 4.14 – Résultat cartographique avant et après application du transport optimal pour un transfert spatial entre la tuile du Sud-Ouest (31TCJ) vers la tuile du Grand-Est (31UEQ) selon le scénario 3. L'emprise est située dans le département de l'Aisne de la tuile cible (31UEQ). Les zones hors forêt ont été masquées (en blanc) avec le masque *High Resolution Layer* (HRL) de 2018.

l'amélioration des performances de la classe peuplier. Nous avons ainsi opté pour le scénario considérant deux classes, une classe peuplier et une classe non-peuplier qui regroupe toutes

les essences pures disponibles dans chacune des tuiles, qui s’est révélé plus pertinent pour l’identification des peupleraies. Nous avons observé que l’amélioration des performances était liée à la nature des couples optimaux formés entre la source et la cible et que l’introduction d’information *a priori* incluant les étiquettes de la source mais aussi de la cible a contribué à l’amélioration de l’identification des peupleraies. Si les résultats statistiques étaient de plus en plus performants avec l’inclusion progressive des termes de régularisations par l’entropie et par les étiquettes, les premières interprétations des résultats cartographiques l’étaient moins avec une sur-détection des peupleraies. Ceci est probablement dû à l’hétérogénéité de la classe non-peuplier et la différence de sa composition entre la source et la cible. L’incorporation des étiquettes de la cible a permis de réduire ces sur-détections mais il serait également envisageable de ne considérer au sein de la classe non-peuplier que les essences pures communes aux deux domaines afin de limiter la formation de couplages incohérents.

Dans les deux sous-sections suivantes, l’objectif est d’évaluer la contribution du transport optimal dans le cas d’un transfert temporel ou spatio-temporel entre deux tuiles Sentinel-2.

4.2.3 Transfert temporel

Le transport optimal a été appliqué dans le cas d’un transfert temporel entre deux années de séries temporelles pour la même tuile. L’objectif étant d’évaluer la capacité du transport optimal à compenser la non-stationnarité de la réponse spectrale d’une année à l’autre et d’utiliser donc un modèle unique entraîné sur une année pour prédire les autres années. Ainsi, il ne serait pas nécessaire de collecter des données de référence pour chaque année, mais plutôt d’adapter la distribution des données sources avec le transport optimal.

La tuile du Sud-Ouest de 2017 est conservée comme source et nous avons utilisé les échantillons de 2018 comme jeu de données du domaine cible. Nous avons considéré les deux classes du scénario 3 à savoir, la classe peuplier avec tous les échantillons d’entraînement disponibles et la classe non-peuplier avec le même nombre d’échantillons répartis entre les classes pures présentes. Comme montré dans le Tableau 4.4, dans la tuile cible du Sud-Ouest de 2018, les mêmes classes ont pu être conservées avec les mêmes proportions que dans la source.

TABLEAU 4.4 – Nombre d’échantillons par classe considéré pour le transfert temporel.

| Transfert | Domaine | Peuplier | Robinier | Châtaignier | Chêne | Hêtre |
|------------------|----------------|----------|----------|-------------|-------|-------|
| Temporel | Source (2017) | 3850 | 1283 | 1283 | 1283 | - |
| | Cible (2018) | 2900 | 967 | 967 | 967 | - |

S’agissant d’un transfert temporel, le but était de conserver exactement les mêmes pixels sources aux mêmes positions et de récupérer leurs valeurs spectrales dans la série temporelle cible de 2018. Compte tenu des contraintes d’acquisition et du changement photo-interprété

dans les peupleraies (principalement des coupes) influençant la phase de nettoyage des données de référence, il existe un écart dans le nombre et la nature des échantillons entre les séries temporelles de 2017 et de 2018 de la même tuile. Bien que toutes les classes aient pu être conservées, nous n'avons pas pu garder les mêmes échantillons aux mêmes positions entre les deux années lors de ce transfert temporel.

Les résultats du transfert temporel avec les quatre variantes du transport optimal sont récapitulés dans le Tableau 4.5. Avant l'application du transport (colonne t_0), les performances atteintes sont de 96% pour l'OA, et de 93% et 99% pour les valeurs de PA des classes peuplier et non-peuplier respectivement. Ceci montre que l'adaptation des distributions n'est pas forcément nécessaire dans ce cas puisque les performances sont déjà très élevées. Lorsque nous avons examiné les profils temporels de la source avant transport et de la cible pour la bande 4 du visible, la bande 5 du proche infrarouge et la bande 11 du moyen infrarouge, les signatures se sont avérées très similaires avec un chevauchement considérable (Figure 4.15).

L'application du transport avec les configurations non régularisé et régularisé a amélioré de 4 à 6% le PA de la classe peuplier qui a atteint sa valeur maximale de 99% avec les deux dernières configurations impliquant les étiquettes de la source et de la cible. Toutefois, on note des fluctuations de -2 à -4% pour la deuxième classe non-peuplier. L'examen des matrices de couplages optimaux montre que la grande majorité des couplages ont été formé entre des échantillons sources et cibles de la même classe avec les coefficients les plus élevés (associés aux coûts les plus faibles) dans les quatre variantes du transport ce qui explique les bonnes performances.

TABLEAU 4.5 – Résultats du transfert temporel avec les quatre configurations de transport optimal selon le scénario 3 des classes. La colonne t_0 indique les performances d'une classification supervisée à partir des données sources non transportées. Les performances sont données en termes de précision globale (OA) et de précision producteur par classe (PA).

| Transfert | Performance | t_0 | EMD | Sinkhorn | Sinkhorn + y^s | Sinkhorn + y^s + y^c ⁽¹⁾ |
|-----------|-----------------|-------|-----|----------|---------------------|--|
| Temporel | OA | 96 | 94 | 98 | 97 | 98 |
| | PA peuplier | 93 | 93 | 97 | 99 | 99 |
| | PA non-peuplier | 99 | 95 | 99 | 95 | 97 |

⁽¹⁾ 100% des étiquettes cibles disponibles.

Nous avons également examiné le résultat cartographique avant et après transport optimal selon les quatre variantes. Dans le premier extrait de la Figure 4.17, nous avons noté un résultat cartographique globalement cohérent avec les performances statistiques. Bien qu'il y ait une amélioration de 4% de la valeur du peuplier entre le cas avant transport et celui du transport avec Sinkhorn, la cartographie est plus fidèle à la réalité terrain avant transport (Figure 4.17b). En effet, nous avons observé certaines sous-détections de peupleraies avec l'approche

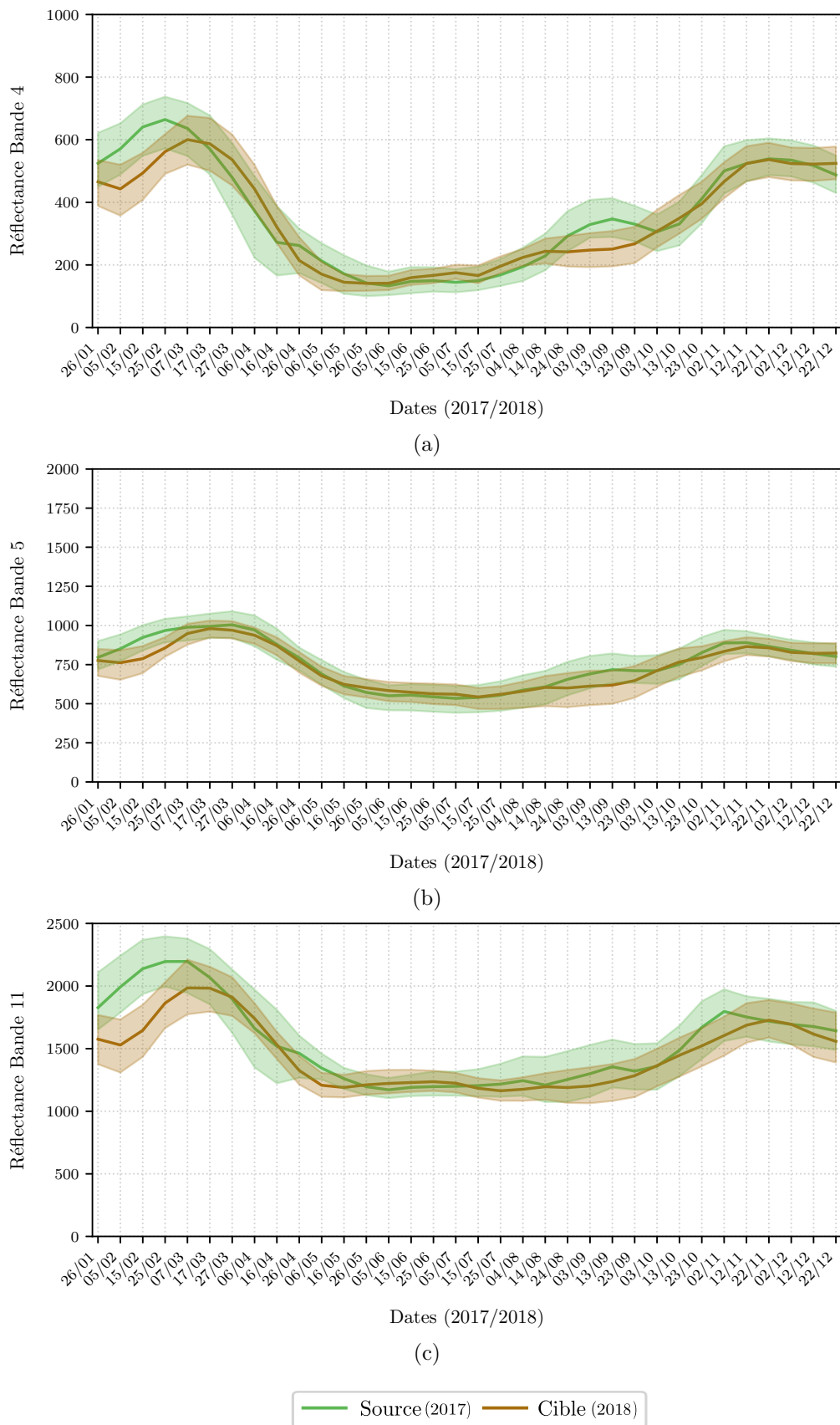


FIGURE 4.15 – Profil temporel de la classe peuplier dans les séries temporelles de la tuile du Sud-Ouest de 2017 (source) et 2018 (cible) avant transport optimal des (a) Bande 4, (b) Bande 5 et (c) Bande 11.

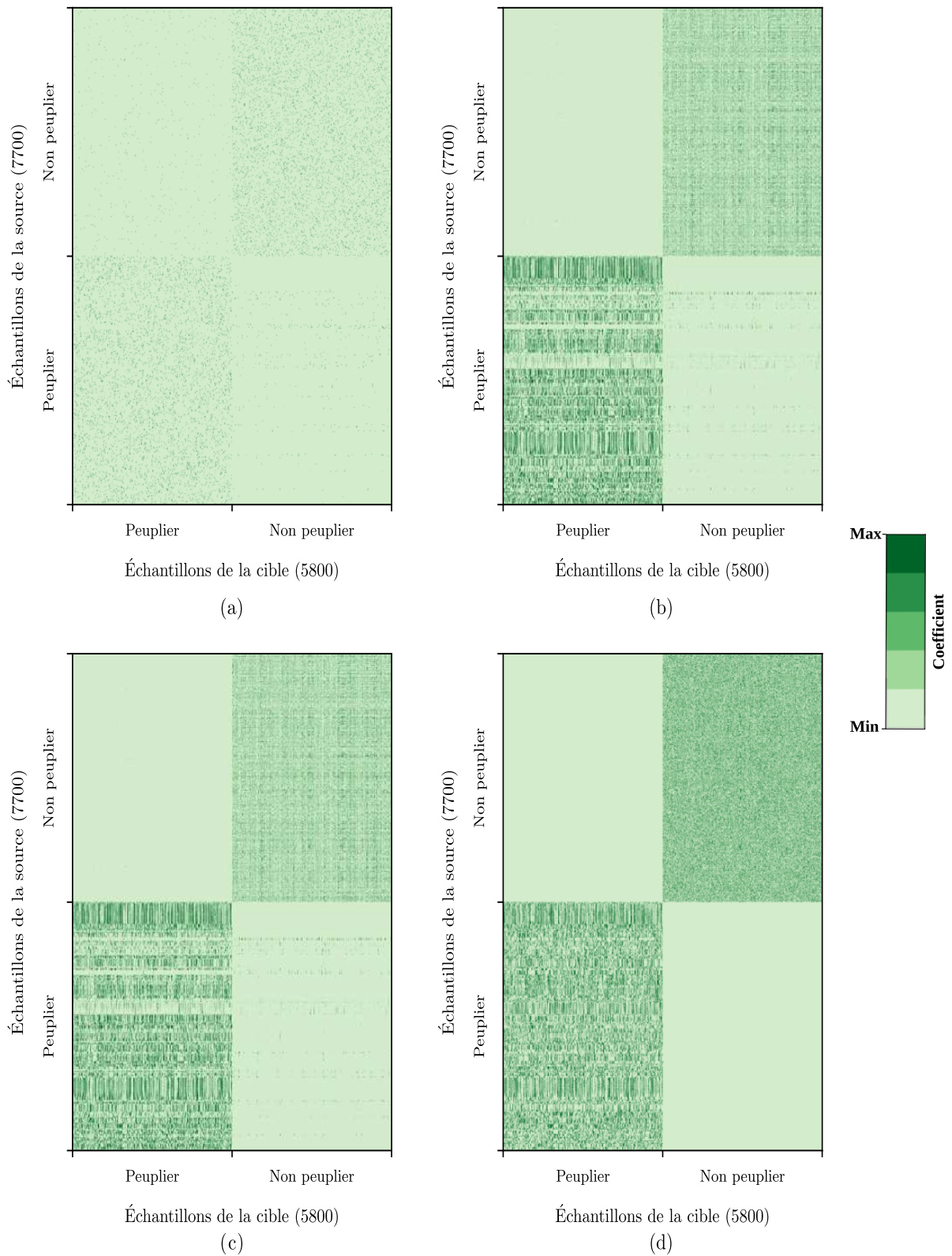


FIGURE 4.16 – Matrices des couplages optimaux entre les échantillons de la sources (lignes) et les échantillons de la cible (colonnes) dans le cas du transfert temporel lors le scénario 3 avec (a) EMD, (b) Sinkhorn, (c) Sinkhorn + y^s et (d) Sinkhorn + $y^s + y^c$.

Sinkhorn de la Figure 4.17d qui étaient pourtant bien identifiées sans adaptation de domaine. Nous avons également remarqué que certaines sur-détections de peupleraies avant transport (Figure 4.17b) étaient tout à fait justifiées et correspondaient à de jeunes plantations qui ne ressortaient pas dans les cartographies après application du transport optimal.

À l'inverse, dans le deuxième extrait de la Figure 4.18, le résultat avant transport a révélé des sur-détections de peupleraies dans des parcelles de chênes purs. Ces confusions ont été rectifiées avec l'application du transport optimal et ont disparu des quatre cartographies (Figures 4.18c, d, e et f).

4.2.4 Transfert spatio-temporel

Dans ce dernier mode de transfert, l'objectif était de tester la capacité du transport optimal à adapter les distributions entre deux séries temporelles localisées sur deux tuiles spatialement distinctes et acquises lors de deux années différentes. Il s'agit de combiner les deux contraintes des deux modes de transfert précédents liées à la variabilité spatiale de la réponse spectrale et à sa variabilité temporelle d'une année sur l'autre.

TABLEAU 4.6 – Nombre d'échantillons par classe considéré pour le transfert spatio-temporel.

| Transfert | Domaine | Peuplier | Robinier | Châtaignier | Chêne | Hêtre |
|-----------------|---------------------------|----------|----------|-------------|-------|-------|
| Spatio-temporel | Source (Tuile 31TCJ 2017) | 3850 | 1283 | 1283 | 1283 | - |
| | Cible (Tuile 31UEQ 2018) | 3000 | - | - | 1500 | 1500 |

De même que pour les deux cas précédents, la source est représentée par les échantillons de la tuile du Sud-Ouest (31TCJ) de l'année 2017. La cible est constituée par les échantillons de la tuile du Grand-Est (31UEQ) de 2018. Le nombre d'échantillons et les classes sont représentés dans le Tableau 4.6. Pour la cible, la classe non-peuplier comporte une seule essence commune avec la source (la classe des chênes) étant donné que la classe des robiniers a été éliminée lors de l'étape de nettoyage des données. Comme détaillé dans 2.4.2, sur les cinq parcelles de robiniers disponibles, les pixels appartenant à trois parcelles d'entre elles correspondaient à des nuages mal-détectés à certaines dates de l'année. En revanche, ayant opté pour un échantillonnage équilibré avec le même nombre de pixels par classe d'une même tuile, le nombre de pixels de robiniers retenus après élimination des valeurs aberrantes était trop faible pour être pris en compte pour toutes les autres classes de la tuile, ce qui nous a conduit à retirer cette classe de l'échantillonnage.

Les résultats avant et après transport optimal pour un transfert spatio-temporel sont indiqués dans le Tableau 4.7. L'application du transport optimal avec l'approche EMD a entraîné une dégradation de -11% de l'OA et de -4% et -19% pour les valeurs de PA des classes peuplier et non-peuplier par rapport aux valeurs atteintes avant l'adaptation de domaine

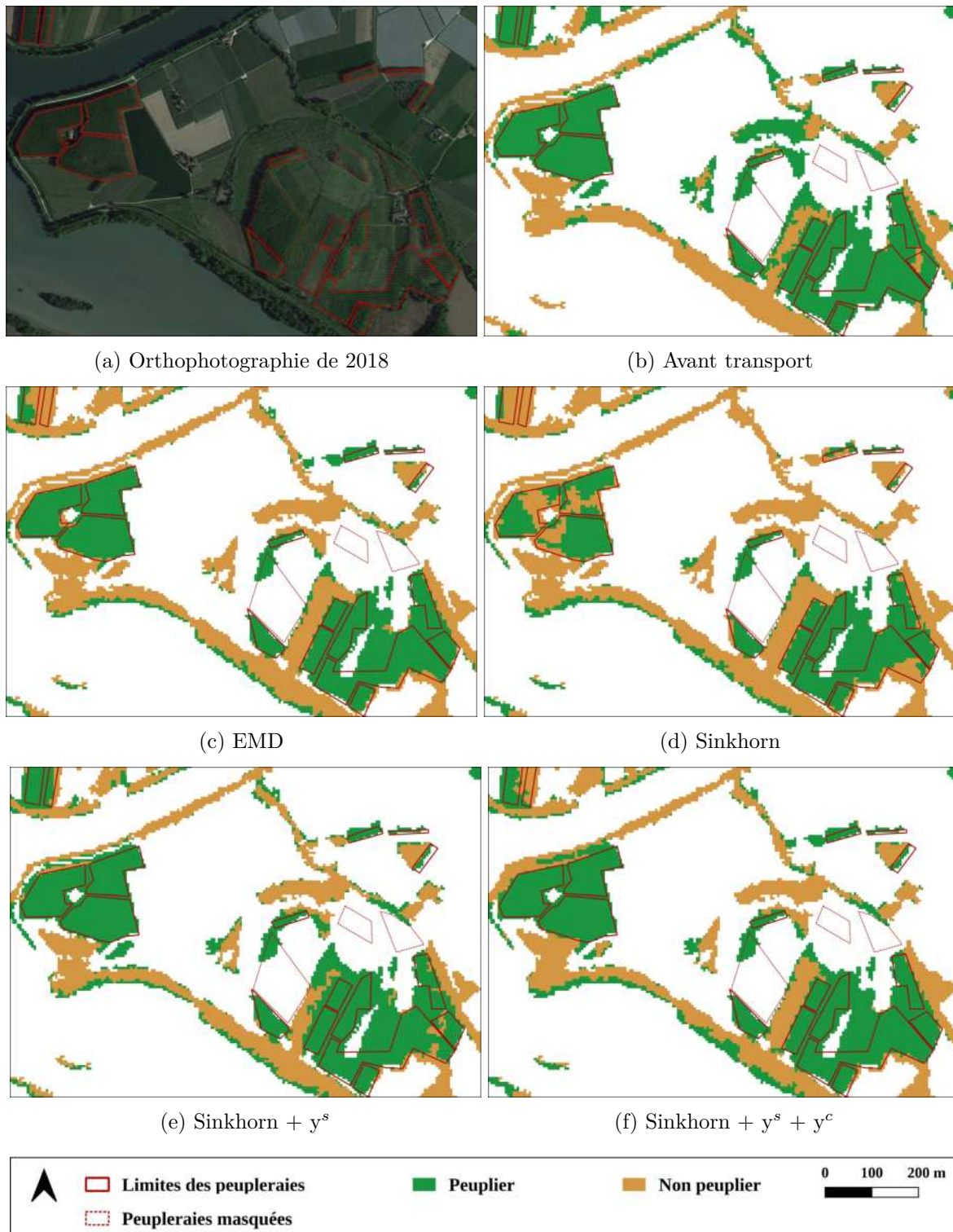


FIGURE 4.17 – Extrait 1 du résultat cartographique avant et après application du transport optimal pour un transfert temporel de la tuile du Sud-Ouest (31TCJ) de 2017 à 2018 selon le scénario 3. L'emprise est située dans le département du Tarn-et-Garonne. Les zones hors forêt ont été masquées (en blanc) avec le masque HRL de 2018.

(colonne t_0). En considérant Sinkhorn, les performances se sont améliorées de +12% en termes d'OA et de +19% et +5% pour le PA des peupliers et des non-peupliers en comparaison avec

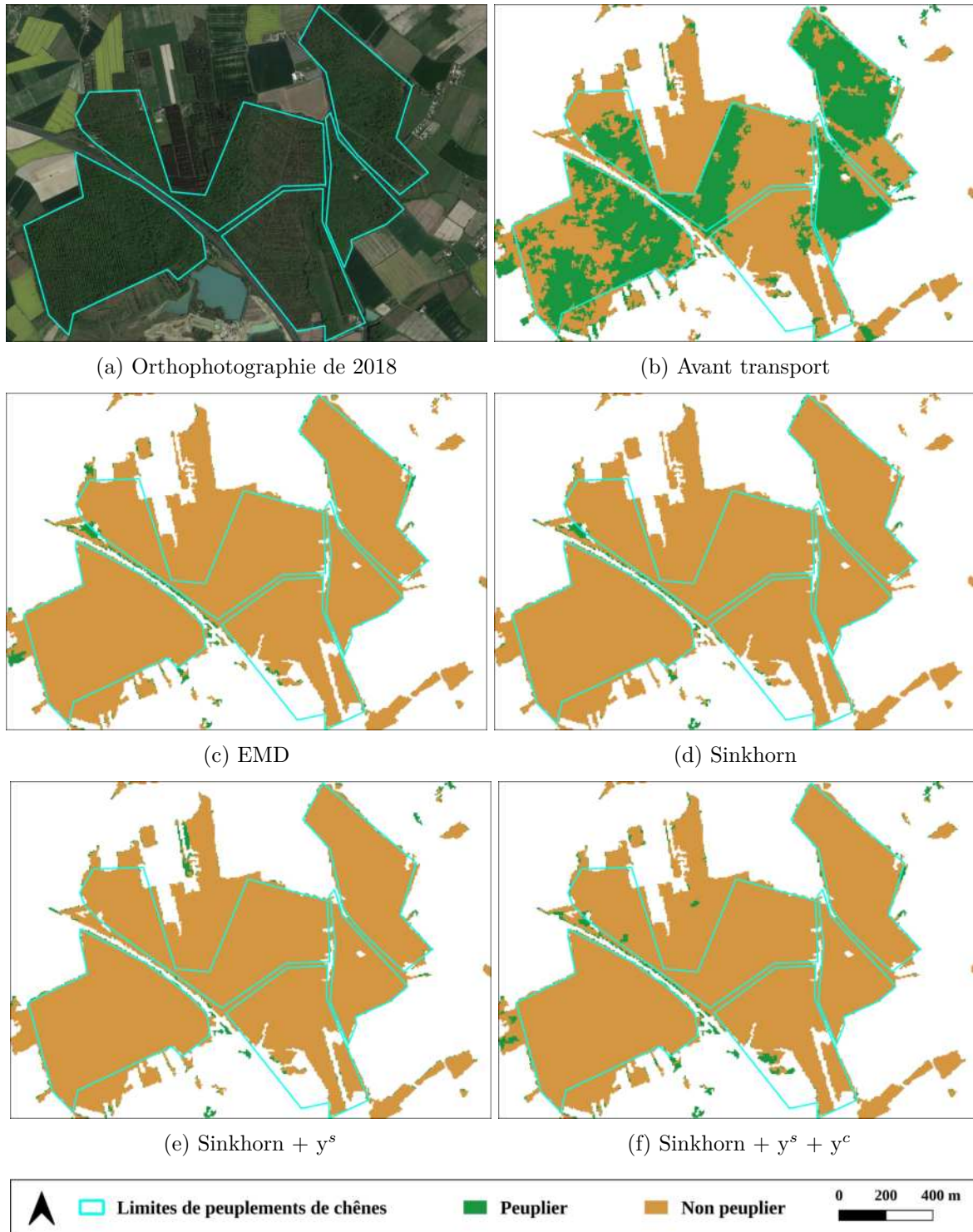


FIGURE 4.18 – Extrait 2 du résultat cartographique avant et après application du transport optimal pour un transfert temporel de la tuile du Sud-Ouest (31TCJ) de 2017 à 2018 selon le scénario 3. L'emprise est située dans le département du Tarn-et-Garonne. Les zones hors forêt ont été masquées (en blanc) avec le masque HRL de 2018.

l'approche précédente. L'ajout des étiquettes de la cible (Sinkhorn + y^s) a légèrement impacté les performances avec -1% et +4% pour les PA des classes de peuplier et non-peuplier. Comme pour les deux modes de transfert précédents, les performances maximales du transport optimal

TABLEAU 4.7 – Résultats du transfert spatio-temporel avec les quatre configurations de transport optimal selon le scénario 3 des classes. La colonne t_0 indique les performances d’une classification supervisée à partir des données sources non transportées. Les performances sont données en termes de précision globale (OA) et de précision producteur par classe (PA).

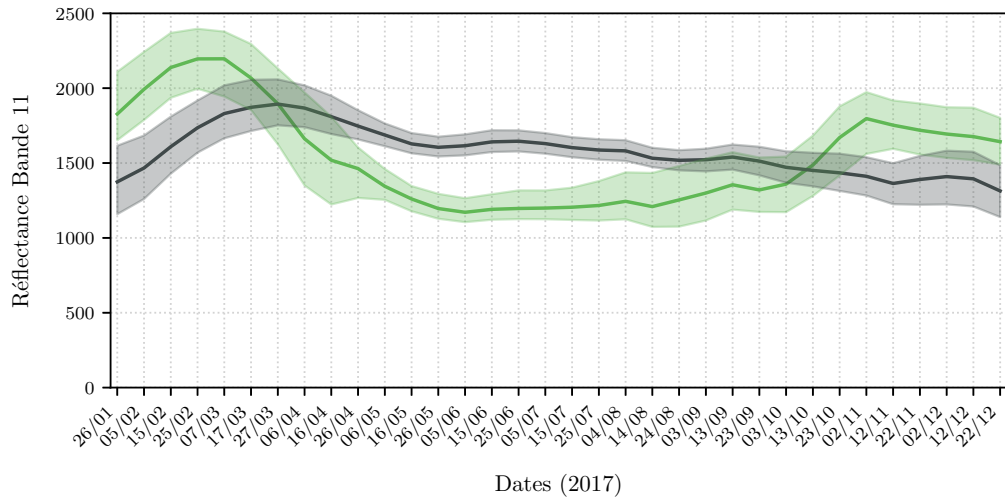
| Transfert | Performance | t_0 | EMD | Sinkhorn | Sinkhorn + y^s | Sinkhorn + y^s + y^c ⁽¹⁾ |
|-----------------|-----------------|-------|-----|----------|---------------------|--|
| | OA | 89 | 78 | 90 | 92 | 96 |
| Spatio-temporel | PA peuplier | 82 | 78 | 97 | 96 | 99 |
| | PA non-peuplier | 96 | 77 | 83 | 87 | 93 |

⁽¹⁾ 100% des étiquettes cibles disponibles.

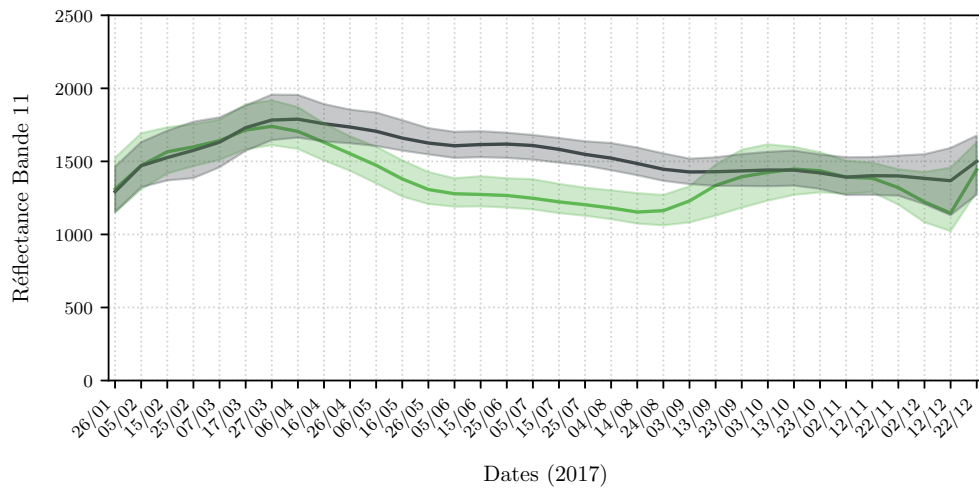
ont été atteintes en incorporant les étiquettes de la source et de la cible (Sinkhorn + y^s + y^c) avec des valeurs de PA de 99% et 93% pour les deux classes de peuplier et non-peuplier.

Dans ce mode de transfert, nous nous attendions à ce que les performances avant transport soient plus faibles que dans le transfert spatial, puisque nous avons en plus fait varier l’année d’acquisition des séries temporelles. Étonnamment, avant transport (t_0), la valeur de PA peuplier était 8% plus élevée et celle de l’OA a augmenté de 2% par rapport au cas de transfert spatial entre les deux mêmes tuiles de la même année. Les échantillons sources étant exactement les mêmes, cette différence peut s’expliquer soit par une meilleure concordance des distributions entre les peupliers des deux domaines puisque les échantillons ne sont pas les mêmes dans la cible entre 2017 et 2018, mais aussi par l’absence des échantillons de robinier de la classe non-peuplier. Comme le PA est indépendant de la prévalence, la différence du nombre d’échantillons ne peut pas a priori en être la raison. Nous avons ainsi examiné les profils temporels des deux classes peuplier et non-peuplier dans la source (tuile 31TCJ de 2017) et dans les deux tuiles cibles du mode spatial (tuile 31TCJ de 2018) et spatio-temporel (tuile 31UEQ de 2018) avant l’application du transport optimal (Figure 4.19). Comme illustré dans la Figure 4.19b, dans la tuile cible du transfert spatial (tuile 31UEQ de 2017), la signature temporelle de la classe non-peuplier dans l’exemple de la bande 11 se chevauche davantage avec celle des peupliers par rapport à la tuile cible du transfert spatio-temporel (tuile 31UEQ de 2018) qui est mieux différenciée, ce qui peut justifier également une meilleure performance dans le cas spatio-temporel.

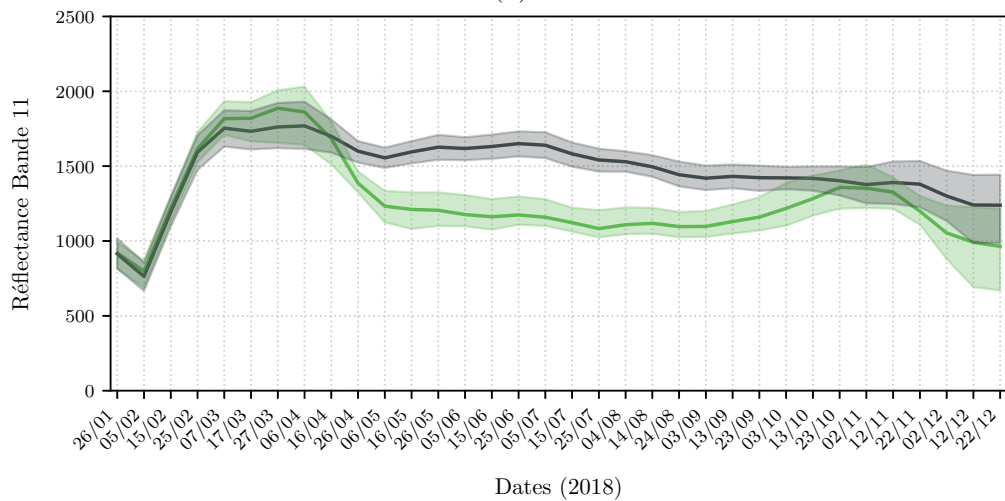
Afin de mieux comprendre les performances des quatre variantes du transport, nous avons tracé les matrices des couplages optimaux (Figure 4.20). Dans l’approche de transport optimal avec l’EMD, nous avons observé que, bien qu’une grande partie des couples optimaux soient formées avec des échantillons sources et cibles appartenant aux mêmes classes, il existe également un nombre non négligeable de couples associant des échantillons de non-peupliers de la source à des échantillons de peupliers de la cible (Figure 4.20a), impactant ainsi les performances. L’augmentation du nombre de liens entre la source et la cible avec l’approche



(a)



(b)



(c)



FIGURE 4.19 – Profil temporel des deux classes peuplier et non-peuplier avant transport optimal avec la bande 11 dans (a) la source (tuile 31TCJ de 2017), (b) dans la cible du transfert spatial (tuile 31UEQ de 2017) et (c) dans la cible dans le cas du transfert spatio-temporel (tuile 31UEQ de 2018).

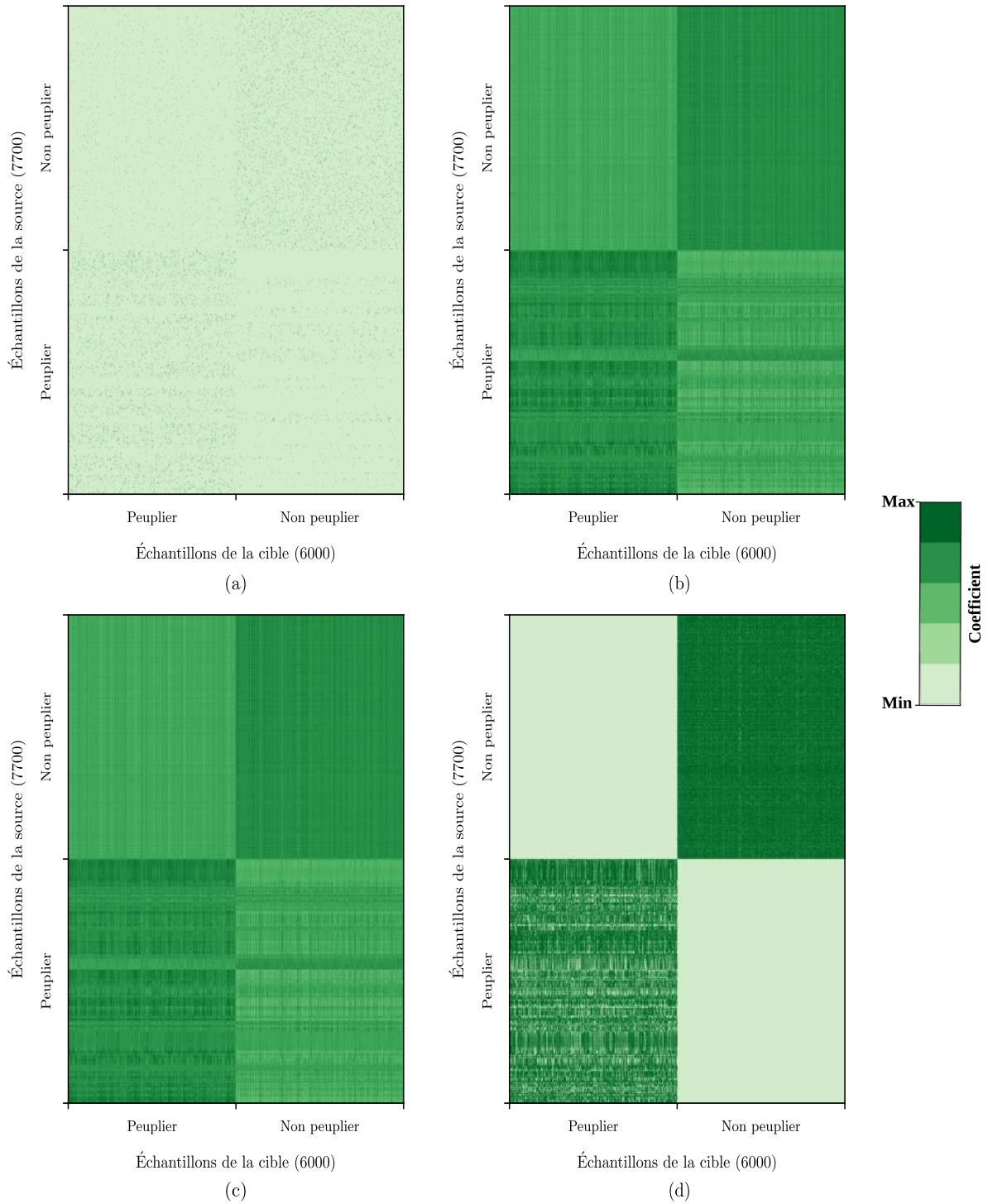


FIGURE 4.20 – Matrices des couplages optimaux entre les échantillons de la sources (lignes) et les échantillons de la cible (colonnes) dans le cas du transfert spatio-temporel lors le scénario 3 avec **(a)** EMD, **(b)** Sinkhorn, **(c)** Sinkhorn + y^s et **(d)** Sinkhorn + y^s + y^c .

Sinkhorn, a conduit à la formation d'un très grand nombre de couplages optimaux (coefficients élevés affichés en vert foncé), impliquant pour la plupart des échantillons sources et cibles de la même classe. Il est important de noter que la valeur optimale du paramètre λ de régularisation par l'entropie retenue pour cette approche est égale à 10^3 ce qui explique le

grand nombre de couples formés entre la source et la cible. Malgré la présence de couplages entre des échantillons sources et cibles n'appartenant pas à la même classe, les performances ont été peu impactées notamment pour la classe des peupliers.

L'ajout de l'information des étiquettes de la source (Sinkhorn $+y^s$) ne semble pas pénaliser de nombreux couples et la matrice des couplages optimaux (Figure 4.20c) est quasiment identique à celle obtenue avec Sinkhorn (Figure 4.20b). Cela explique ainsi la faible variation des performances entre les deux approches.

La dernière approche, qui intègre les étiquettes de la source et de la cible, a permis d'éliminer tous les couplages associant des échantillons sources et cibles n'appartenant pas aux mêmes classes, atteignant ainsi les meilleures performances.

Dans cette section, le transfert a été testé à partir d'une seule source, la tuile source 2017 du Sud-Ouest, et ce pour les trois modes de transfert spatial, temporel et spatio-temporel. Dans ce travail préliminaire des approches d'adaptation de domaine, nous n'avons pas évalué la contribution du transport optimal depuis d'autres sources, ce qui pourrait conduire à des résultats différents. Cette question sera étudiée plus en détail dans le deuxième chapitre de cette partie.

Il est important de souligner que tous les résultats rapportés sont obtenus avec les paramètres optimaux du classifieur (définis dans la Section 3.2.2) et du transport optimal (λ et η) dans le cas des configurations régularisées. Nous avons observé de très grands écarts de performance avec de faibles variations de λ et η , ce qui rend l'étape d'optimisation essentielle pour le transport optimal.

4.3 Conclusion

Dans cette première exploration des approches d'adaptation de domaine par adaptation des distributions, le transport optimal non régularisé a montré son potentiel pour la classe peuplier mais également pour la plupart des classes d'essences pures quand elles sont considérées individuellement.

L'inclusion du terme de régularisation par l'entropie permettant de se baser sur plusieurs échantillons de la cible pour déplacer ceux de la source s'est avérée utile dans les trois modes de transfert et lorsque deux classes peuplier et non-peuplier ont été considérées. Dès que le nombre de classes augmente ou que des classes hétérogènes sont considérées, la densité des couplages optimaux avec les classes proches augmente et les performances se dégradent. Outre la ressemblance entre les classes, la présence de bruit dans les échantillons provenant de valeurs aberrantes ou de changements, qui n'ont pas été détectés pour des classes autres que le peuplier, peut également avoir un impact sur la qualité du transport. Ces résultats

confirment ce qui a été souligné par (Tardy et al., 2019) où le choix des échantillons a un impact significatif sur la qualité du transport.

Du point de vue du transport optimal, cela met en évidence la complexité de l'espace des caractéristiques représenté par 340 variables (10 bandes spectrales \times 34 dates) qui, au-delà des exigences en termes de calcul, peut rendre difficile l'estimation des couplages optimaux. Ce point a été souligné par Courty et al. (2014), où la présence de valeurs aberrantes ainsi que la haute dimensionnalité des données source et cible peuvent entraîner des irrégularités dans la fonction de transport. Par ailleurs, la fonction de coût considérée est la distance Euclidienne qui s'avère moins appropriée pour les grandes dimensions (Cassisi et al., 2012). D'autres mesures de distance telle que la déformation temporelle dynamique ou *Dynamic Time Warping* (DTW) plus adaptées aux séries temporelles (Cuturi and Blondel, 2017) pourraient être envisagées.

Nous avons également montré qu'il est essentiel d'avoir les mêmes classes principalement dans les approches de transport optimal non supervisé et sans information sur les étiquettes pour éviter la formation de couplages incohérents et donc la dégradation des performances. Même si nous avons considéré une classe non-peuplier avec tous les échantillons d'essences pures disponibles, il est aussi important que la composition soit cohérente entre les domaines source et la cible afin de ne pas ralentir l'amélioration des performances. Ceci constitue une réelle contrainte qui n'est pas forcément possible à chaque fois, surtout si les proportions de classes entre les deux domaines doivent rester équivalentes.

Sans information sur les étiquettes, la présence des classes hétérogènes a tendance à ralentir l'évolution des performances mais dès que ces informations *a priori* sont introduites, l'amélioration est nette. Nous nous sommes donc passés d'une approche non supervisée vers une approche semi-supervisée impliquant une mise à disposition de données cibles étiquetées ce qui nous renvoie à la problématique initiale, à savoir la nécessité de disposer d'échantillons étiquetés dans chaque nouvelle zone d'étude spatialement ou temporellement distincte. Bien que la quantité des étiquettes cibles optimale était autour de 50% du nombre total d'étiquettes disponibles, cela représente un nombre important d'étiquettes qui peuvent être difficiles à collecter. D'autre part, les performances obtenues avec ce nombre doivent être interprétées avec précaution car elles sont fortement liées à la sélection aléatoire unique effectuée sur les échantillons. Ainsi, il serait pertinent de répéter la sélection aléatoire afin de dériver des performances moyennes ou encore de disposer d'un critère de sélection des étiquettes cibles à incorporer qui permettrait d'orienter le choix et d'analyser leur nature en termes de classe d'appartenance et pas seulement en termes de nombre.

Cela nous oriente vers d'autres approches d'adaptation de domaine, notamment l'apprentissage actif qui permet de rajouter un nombre limité d'échantillons étiquetés de la cible en

fonction de leur pouvoir informatif et de leur capacité à mieux détecter la cible. Cette approche sera abordée dans le chapitre suivant.

D'un point de vue cartographique, les performances statistiques n'étaient pas toujours en accord avec le résultat cartographique. Le résultat sans adaptation des distributions était dans certains cas plus fidèle à la réalité terrain que celui obtenu après application du transport optimal. Ce point mérite d'être approfondi par une vérification sur le terrain et une analyse cartographique plus approfondie avec un protocole de validation approprié.

4.4 Bilan

Malgré le potentiel du transport optimal, les contraintes liées au choix des échantillons et des classes de départ pour assurer l'estimation des couplages, le besoin d'étiquettes de la cible pour garantir les meilleurs résultats ainsi que la complexité du calcul qui n'est pas seulement liée à la grande dimension des données mais aussi à la nécessité d'optimiser finement les paramètres du transport, font de cette méthode un défi pour une application sur de grandes étendues. Une autre contrainte peut se présenter et concerne le cas où plusieurs cibles sont à considérer. Cela va conduire à une adaptation des distributions de la source pour chacune des cibles, ce qui peut être à la fois chronophage mais aussi contradictoire si les deux cibles sont très différentes. Enfin, la dernière contrainte du transport optimal et la plus importante à notre sens, réside dans la complexité de l'approche pour son transfert à des partenaires forestiers non spécialistes, limitant ainsi la possibilité d'une mise en œuvre opérationnelle.

Cette première étude exploratoire sur l'utilisation de l'adaptation de domaine pour la cartographie des peupleraies sur de grandes étendues constitue un travail préliminaire qui nous a orienté vers d'autres choix méthodologiques à savoir, de retenir l'adaptation de domaine comme une méthode de passage à l'échelle mais de définir une approche plus simple à mettre en œuvre sur de grandes étendues de manière plus opérationnelle.

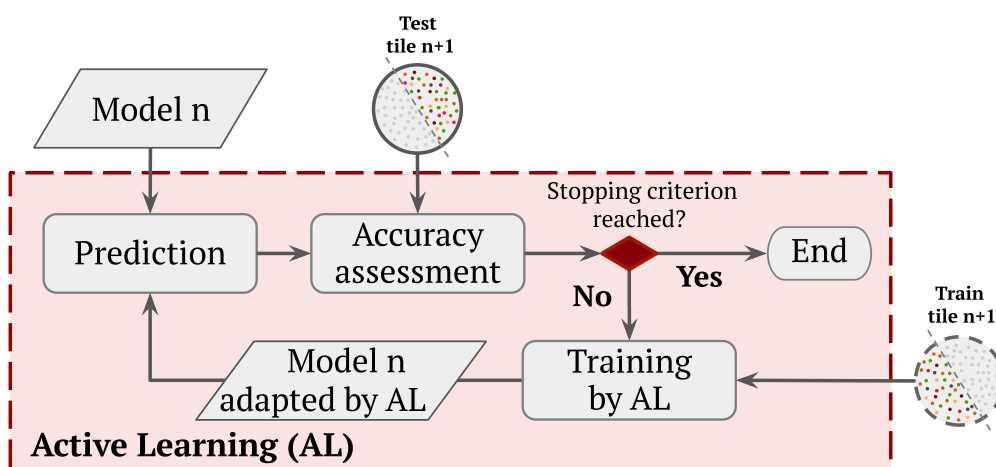


Chapitre 5

Adaptation de domaine par apprentissage actif

Sommaire

| | | |
|-----|---|-----|
| 5.1 | Résumé | 111 |
| 5.2 | From local to global : A transfer learning-based approach for mapping poplar plantations at national scale using Sentinel-2 | 112 |
| 1. | Introduction | 113 |
| 2. | Materials | 117 |
| 3. | Methods | 120 |
| 4. | Results | 127 |
| 5. | Discussion | 135 |
| 6. | Conclusion | 142 |
| 5.3 | Bilan | 144 |



5.1 Résumé

Dans ce chapitre, nous proposons une deuxième approche d'adaptation de domaine pour répondre aux contraintes de la large échelle.

Bien que les méthodes de classification supervisée se soient avérées très précises pour cartographier le couvert forestier à partir d'images de télédétection, leurs performances dépendent en grande partie des échantillons étiquetés utilisés pour construire les modèles. En plus de leur coût d'acquisition élevé, ces échantillons sont souvent rares et ne sont pas entièrement représentatifs de la variabilité des distributions des classes. Par conséquent, lorsque les modèles de classification sont appliqués sur de grandes étendues présentant une forte variance intra-classe, les performances s'effondrent généralement en raison des problèmes de décalage des données (*datashift*).

Dans ce chapitre, nous proposons l'utilisation de l'apprentissage actif pour adapter efficacement un classifieur entraîné sur une image source, à des images cibles spatialement distinctes avec un minimum d'échantillons étiquetés sans pour autant sacrifier la performance de la classification. L'adaptation consiste à enrichir en cascade un modèle local initial par de nouveaux échantillons d'entraînement pertinents provenant d'autres zones et qui améliorent de manière itérative les capacités de généralisation du classifieur afin d'aboutir à un modèle global adapté à ces différentes zones. Cette sélection active s'appuie sur l'échantillonnage d'incertitude pour se concentrer directement sur les pixels les plus informatifs pour lesquels l'algorithme est le moins certain de leurs étiquettes de classe. Les expériences menées sur les séries temporelles Sentinel-2 ont révélé leur grande capacité à identifier les plantations de peupliers à l'échelle locale avec un F-score moyen allant de 89,5% à 99,3%.

Pour l'adaptation sur de grandes étendues, les résultats ont montré que lorsque le même nombre d'échantillons d'entraînement était utilisé, l'apprentissage actif surpassait l'échantillonnage aléatoire jusqu'à 5% de la précision globale et jusqu'à 12% du F-score de la classe. En outre, et selon la classe considérée, le modèle d'échantillonnage aléatoire a nécessité jusqu'à 50% d'échantillons supplémentaires pour obtenir les mêmes performances qu'un modèle basé sur l'apprentissage actif. En outre, les résultats démontrent l'aptitude du modèle global dérivé à cartographier avec précision les plantations de peupliers parmi d'autres espèces d'arbres, avec des valeurs de précision globale jusqu'à 14% supérieures à celles obtenues avec les modèles locaux. L'approche proposée ouvre la voie à une cartographie à l'échelle nationale dans un contexte opérationnel.

Mots-clés : Apprentissage actif ; Adaptation de domaine ; Peupleraies ; Transfert spatial ; Grandes étendues ; Sentinel-2

5.2 From local to global: A transfer learning-based approach for mapping poplar plantations at national scale using Sentinel-2

Authors: Yousra Hamrouni, Eric Paillassa, Véronique Chéret, Claude Monteil, David Sheeren

Paper status: Published in ISPRS Journal of Photogrammetry and Remote Sensing on 20 November 2020: <https://doi.org/10.1016/j.isprsjprs.2020.10.018>.

Abstract: Reliable estimates of poplar plantations area are not available at the French national scale due to the unsuitability and low update rate of existing forest databases for this short-rotation species. While supervised classification methods have been shown to be highly accurate in mapping forest cover from remotely sensed images, their performance depends to a great extent on the labelled samples used to build the models. In addition to their high acquisition cost, such samples are often scarce and not fully representative of the variability in class distributions. Consequently, when classification models are applied to large areas with high intra-class variance, they generally yield poor accuracies because of data shift issues. In this paper, we propose the use of active learning to efficiently adapt a classifier trained on a source image to spatially distinct target images with minimal labelling effort and without sacrificing the classification performance. The adaptation consists in actively adding to the initial local model new relevant training samples from other areas in a cascade that iteratively improves the generalisation capabilities of the classifier leading to a global model tailored to these different areas. This active selection relies on uncertainty sampling to directly focus on the most informative pixels for which the algorithm is the least certain of their class labels. Experiments conducted on Sentinel-2 time series revealed their high capacity to identify poplar plantations at a local scale with an average F-score ranging from 89.5% to 99.3%. For large area adaptation the results showed that when the same number of training samples was used, active learning outperformed random sampling by up to 5% of the overall accuracy and up to 12% of the class F-score. Additionally, and depending on the class considered, the random sampling model required up to 50% more samples to achieve the same performance of an active learning-based model. Moreover, the results demonstrate the suitability of the derived global model to accurately map poplar plantations among other tree species with overall accuracy values up to 14% higher than those obtained with local models. The proposed approach paves the way for a national scale mapping in an operational context.

Keywords: Active learning; Domain adaptation; Poplar plantations; Spatial transfer; Large areas; Sentinel-2

1. Introduction

Poplar (*Populus spp.*) is one of the fast-growing and wood producing trees which are increasingly considered as an important resource to meet the global demand for natural forest products (Isebrands and Richardson, 2013). According to a report from the Food and Agriculture Organization of the United Nations (FAO) at the 25th session of the International Poplar Commission (IPC), France has the largest surface area of planted poplar in Europe with about 0.2 million hectares (FAO, 2016a) corresponding to approximately 10% of the French planted forests.

In France, poplar cultivation is a key local industry which involves silviculturists, nurserymen, forest managers and wood processors. It is coordinated by the National Poplar Council (CNP). Despite the increased demand for poplar wood, planted areas continue to decrease (Robert, 2018). Indeed, for the past 20 years, the poplar sector has undergone several economic, social and environmental upheavals that have consequently had an impact on the planting rate (Paillassa, 2014). This deficit of several years has led the sector to an inevitable wood shortage, which is expected to reach at least 500,000 m³/year in 2025, according to the CNP. Considering this situation, national strategies have been undertaken to encourage all the industry stakeholders, including providing financial incentives to replant poplar.

As the future of poplar in France relies on these replanted areas, obtaining spatially explicit information on newly-planted and cleared surfaces is crucial. This requires precise mapping and timely characterisation of their spatial and temporal dynamics. However, accurate and regularly updated maps of poplar plantations are still not available at the national scale and the update rate of existing French forest databases is not suitable for this species because of its short rotation cycle (15 years on average).

Since the availability of high spatial, spectral and temporal resolution imagery from the Copernicus Sentinel sensors, new opportunities for monitoring poplar plantations over large areas have emerged. Several works have already demonstrated the potential of remote sensing for mapping plantations such as rubber (Li and Fox, 2011; Dong et al., 2012; Ye et al., 2018; Han et al., 2018; Xiao et al., 2019; Poortinga et al., 2019) and palm oil (Rosenqvist, 1996; Lazecky et al., 2018; Cheng et al., 2018; Poortinga et al., 2019; Descals et al., 2019) but only a few have focused on poplar (Chardenon and Flouzat, 1981; Borry et al., 1993; Heyman et al., 2003; Grignetti et al., 2009; Eslami and Zahedi, 2011). The few studies which did, have mainly mobilised very high spatial resolution data which may be expensive and not necessarily accessible everywhere. Despite the promising results reported, most of these studies were limited to relatively small areas and the performances depended largely on the data used for training and validation. Consequently, their application over large and heterogeneous areas remains challenging due to their limited generalisation ability (Li and Fox, 2012; Cheng et al.,

2018).

Traditionally, classification methods assume that training and test sets share the same feature space and the same distribution (Pan and Yang, 2010). In remote sensing, one model is generally computed per area, including all the variability in the data. The local classification models work well on other areas as long as the distribution has not changed. However, over large areas, spectral reflectance signatures may vary according to acquisition conditions (incidence angle, illumination, relief). Moreover, in the particular case of vegetation and for a given tree species, the observed spectra may shift significantly from one area to another due to several factors including the spatial variability of the phenology (seasonal changes), conditions at the site (the type of soil, moisture conditions, etc.) and the forest health status (Lambert et al., 2013). Direct application of the original and local trained model in new areas (i.e. across-region generalisation (Woodcock et al., 2001)) will therefore result in poor classification performances (Crawford et al., 2013).

Several options could be considered for large area mapping. The first is to build site-specific models which are locally tailored to each study area. Another is to train a global model using all the reference samples of each local site. These usual strategies can work if labelled data for training are available everywhere and are sufficiently representative of the spatial variability. At a national scale, labelled data can easily be retrieved from digital forest maps (if they exist). However, their quality may be unsuitable as they are not frequently updated. These databases often contain noisy data. Hence, the classifications derived from such references may perform poorly because they do not always adequately reflect the reality on the ground.

In order to ensure more accurate and reliable results, representative and updated labelled data have to be collected, which remains the major bottleneck. Field surveys are unavoidable. However, since they are economically and timely expensive, they could only be carried out on limited extents. Manual labelling by human experts can also be considered but only for well-targeted areas. At a national scale, the acquisition of large reference data sets through photo-interpretation is unrealistic. In addition, photo-interpretation may introduce some bias through information redundancy (Tuia et al., 2009; Crawford et al., 2013). Indeed, several neighbouring pixels may contain the same spectral information which can skew the learned model thereby reducing its generalisation ability (overfitting). Thus, methods that help to minimise human expertise while providing accurate classification results over large areas are required.

In view of this, an alternative option would be to build a local classification model on an initial well-informed study area (i.e. containing sufficient and accurate labelled samples for training) and to adapt it to a new area with a limited number of cleaned and up-to-date

samples collected by human experts. From a machine learning point of view, this strategy refers to *transfer learning* in which, a model trained locally using an initial feature space on a *source domain*, is adapted, to accurately predict a new area called the *target domain* based on a different feature space (Pan and Yang, 2010). The adaptation is supposed to minimise the number of extra labelled examples needed to achieve good classification performance.

Transfer learning has been developed over the last 15 years in response to these challenges, by leveraging the knowledge gained during initial learning tasks and adapting it to the subsequent ones. In the literature, transfer learning can be categorised into three branches, depending on the availability of labelled data and the relationship between the source and target domains (Pan and Yang, 2010). When the labelled data are only available in the source domain and the learning task is the same in both source and target domains, the transfer learning branch is referred to as *domain adaptation* (DA). In the remote sensing literature, DA techniques can be grouped into four families (Tuia et al., 2016):

- DA *by selecting invariant features*: this consists in reducing the feature space by selecting only the image features (e.g. raw spectral bands or derived attributes) that are the most robust to change. The latter are described as landmarks in Gong et al. (2013) and have the same distribution in both source and target domains. The new feature space is thus more stable and the gap between the two domains is reduced.
- DA *by adapting data distribution*: the aim of this second approach is to create a common feature space for both domains from the two distributions in order to fit the classifier that remains unchanged. Several strategies can be applied such as feature extraction (Matasci et al., 2015) or similarity-based methods (Petitjean et al., 2011) for data alignment.
- DA *by adapting the classifier*: in this semi-supervised approach, the classification model already built with the source labelled data is modified to fit the target domain by only considering the distribution of the unlabelled data and never their labels. The classifier is therefore updated to accurately classify the target data.
- DA *by active learning*: it is a particular case of the previous approach in which the classifier is allowed to acquire some labelled data from the target domain. These data are selected according to their potential for improving the initial classification model to correctly predict the target domain (Tuia et al., 2011a).

Among these different approaches, DA by active learning is a good candidate to address the issues related to large area mapping using image data. It allows building a global model from a local one by updating the source classifier with a minimum of information from the target domain.

Although active learning (AL) is well-known in the remote sensing community (Tuia et al., 2011b; Crawford et al., 2013), this advanced technique is still underused for real-world and



practical applications. A large number of studies on AL has focused on the most efficient sampling strategy to select informative and representative samples from a single image (i.e. domain), exploiting specific heuristics of the learning algorithms (Mitra et al., 2004; Tuia et al., 2011b) or introducing new ones (Demir et al., 2011; Volpi et al., 2012), including spatial constraints (Pasolli et al., 2014; Stumpf et al., 2014). Spatial heuristics combined with other ranking strategies of samples were defined, in particular, to take the real cost of annotation into account (Demir et al., 2014; Persello et al., 2014; Malek et al., 2019). Sampling strategies were often studied for pixel-based classification but also for object-based image analysis (Amor et al., 2018; Ma et al., 2018). Other adaptations have been proposed to integrate feature extraction simultaneously in the AL process (Pasolli et al., 2016) or to combine multiple feature types (e.g. from multimodal or multisensor data sets) by learning multiple metrics jointly, each metric being devoted to a specific feature type (Zhang et al., 2016). Currently, multiview AL is developing (Hu et al., 2018; Zhang et al., 2020) as AL based on deep learning (Liu et al., 2017; Cao et al., 2020).

In the majority of existing studies, AL is applied to efficiently select a reduced set of optimal training samples needed by classifiers but only rarely in a transfer perspective. This context of application is challenging due to the data shift mentioned previously. This DA problem has been initially addressed with unsupervised or semi-supervised learning techniques and especially, between temporal domains (Bruzzone and Prieto, 2001; Bruzzone and Cossu, 2002). Few papers explored AL for spatial transfer. Learning under covariate shift or sample selection bias has been considered in Rajan et al. (2008); Tuia et al. (2011a) and Persello and Bruzzone (2012). However, the experiments were carried out at a local scale, within the same scene (by defining disjointed areas) or using two images covering two distinct but very close areas. In addition, because these studies mainly focused on methodological aspects of AL for DA, the images used often cover a very small extent. In this case, the data shift between the source and target domains may be limited compared to large and very far geographical areas that include additional variability, due in particular to biogeographical effects. Adaptation by AL for mapping large areas has been investigated in Alajlan et al. (2014). In this work, the authors proposed an AL-based approach using a MODIS data set to generate a land cover map at a continental (European) scale with the possibility to have a class present in a target domain but missed in the source domain. This study is very close to ours. However, using a single subdivided MODIS image, they focused on very general classes (cloud, snow, vegetation, water, other) that do not require the temporal dimension and for which, the spatial non-stationarity may be less pronounced due to the spatial resolution (1 km) reducing heterogeneity.

In this paper, we propose a robust active learning-based strategy for spatial generalisa-



tion to map a wide range of poplar plantations in France using Sentinel-2 time series. The originality of this work is twofold. The first lies in positioning the application of AL for DA in an operational context while considering all its main challenges: the high dimension of hyper-temporal data, the irregularity of satellite image acquisition, the spatial and temporal variability of class reflectances, and the availability of reference data, possibly noisy. To the best of our knowledge, AL for DA has not yet been deployed with image time series for such a real-world application at this scale. In addition, the ability to provide an accurate classification of poplar plantations at such a scale is yet to be demonstrated. The second contribution consists in providing a comprehensive understanding of the functioning of AL in this context.

The proposed approach consists in adapting a local classifier, built on labelled data from a source domain, with a limited number of new and relevant labelled samples from other areas (i.e. target domains) by AL. This cascading update of the initial local model improves its generalisation capabilities and leads to a global model appropriated for a national scale mapping. Beyond the overall accuracy metrics, we evaluate the contribution of AL for the different model classes. Besides, we investigate the impact of the AL transfer direction (from one region to another), the impact of the AL ranking strategy (uncertainty and diversity) to select the most informative samples in the target domain and the behaviour of AL towards missing classes. This approach is compared to a traditional passive classification strategy that builds a global model by randomly selecting new samples from the target domains. We assume that similar classification performances can be reached with both approaches but using much fewer data samples with AL. Local models (i.e. models based on site-specific data only) are also computed with both outdated reference data and updated data by photo-interpretation, for comparison. This analysis aims at demonstrating and quantifying the importance of checking and cleaning training data derived from digital forest maps, supporting the AL approach for large areas since it optimises the human intervention.

2. Materials

2.1 Study areas

Three different study sites were chosen with forest partners. They are representative of the variability of French poplar plantations in terms of cultivars, silvicultural practices and climatic conditions. The poplar sites are located in northeastern, central and southwestern France and are covered by three Sentinel-2 tiles with a surface area of 100 km² each (Figure 5.1).

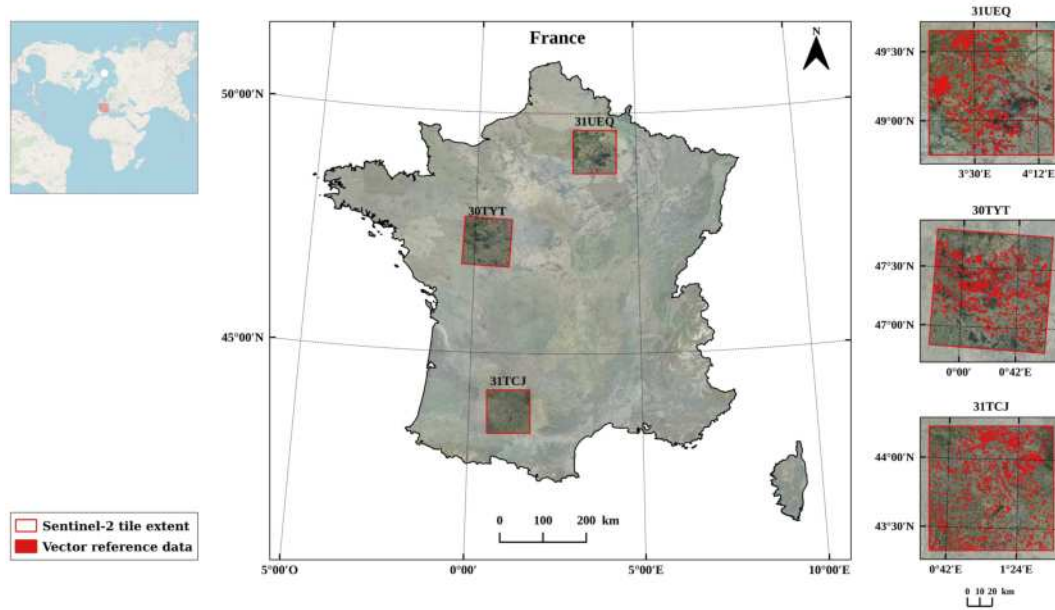


Figure 5.1 – Outline of the study areas used in this work. 31UEQ, 30TYT and 31TCJ are the Sentinel-2 extents at the UTM tiling following the US-MGRS (US Military Grid Reference System). Vector reference data used for each tile are displayed as red polygons.

2.2 Satellite imagery

In this study, Sentinel-2 (S2) optical *Satellite Image Time Series* (SITS) acquired in the year 2017 were used. We selected the 10 S2 bands available at 10 m and 20 m spatial resolutions. The 60 m bands dedicated to atmospheric correction and cloud screening were excluded from the analyses.

All available acquisitions for the three study areas (S2 tiles) were downloaded from the French Theia Land Data Center. They are surface reflectance products (level 2A) processed by the MAJA¹ software, orthorectified, atmospherically corrected and provided with a cloud mask (Hagolle et al., 2015). From one tile to another, the acquired images are not synchronous, either because of sensor-specific constraints or the presence of clouds. The number of downloaded dates per tile with the acquisition period and the importance of cloud coverage is reported in Figure 5.2.

2.3 Reference data

Reference data for training and testing were obtained from the most recent version of the French Forest database (BD Forêt® IGN, v.2) created by the national forest inventory and mapping agency (IGN). The vector data are provided by district and created through photo-interpretation of high-resolution aerial photographs. It took about 11 years to get com-

1. Software developed in coordination between CNES/CESBIO and DLR using the multi-sensor atmospheric correction and cloud screening (MACCS) method.

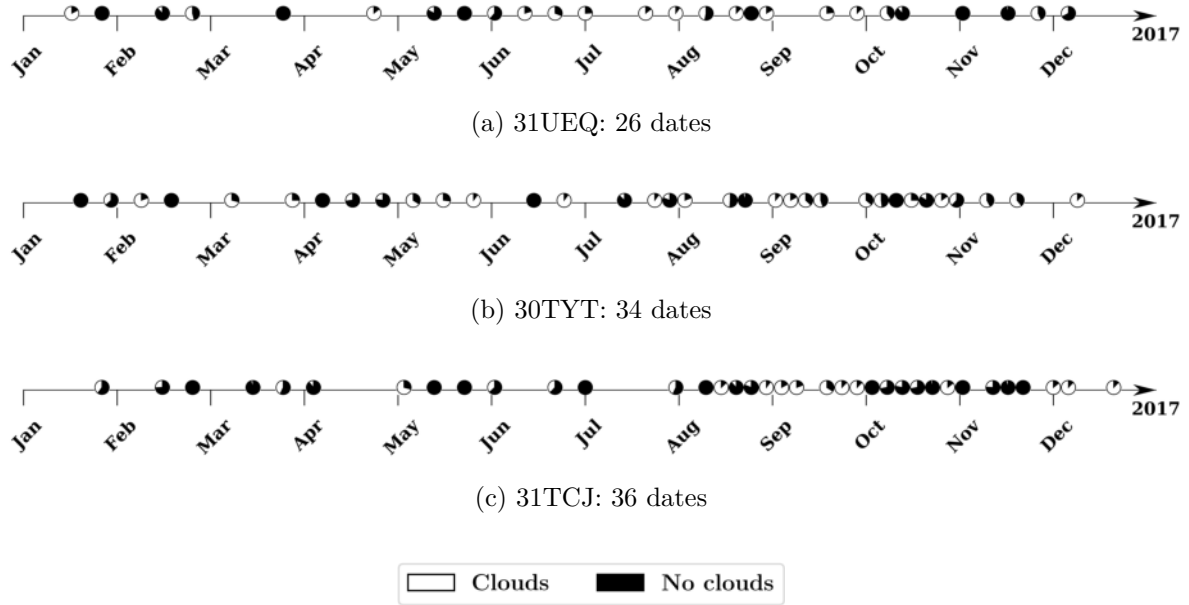


Figure 5.2 – Sentinel-2 acquisitions available in 2017 and percent cloud cover/date in (a) the north-eastern tile 31UEQ, (b) the central tile 30TYT and (c) the south-western tile 31TCJ.

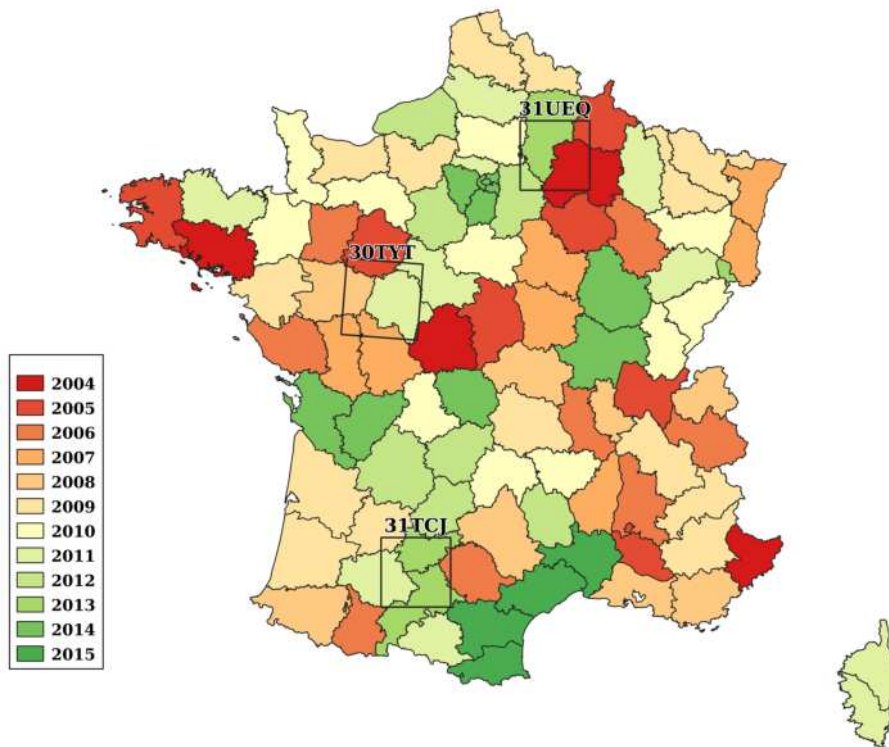


Figure 5.3 – Acquisition years of aerial photographs per district in the forest database used as reference.

plete national coverage with orthophotos (Figure 5.3) and the forest maps generally became available from two to six years later.

In this study, reference data were extracted for all existing deciduous species as polygons of pure and mixed stands with a minimum area of one hectare (Figure 5.1). Pure stands are stands in which at least 75% of the total extent is covered by a single dominant species

while in mixed stands (i.e. polygons) the coverage is between 50% and 75% (IGN, 2014). The polygons include poplar plantations which are always defined as pure stands in the database.

The time lag between the S2 images year (2017), the aerial photographs acquisitions (from 2004 to 2015) and the forest database production period (from 2007 to 2018) could potentially be a source of error in the reference data. With an average production cycle of 15 years, poplar plantations are the forest stands most likely to be affected by this time lag. Consequently, two reference data sets were created. The first one included reference polygons drawn directly from the forest database (outdated data). For the second data set, an additional validation was performed but only on poplars. It consisted in visually inspecting all poplar polygons to check for possible changes and to ensure all references were up to date (updated data). The validation was performed by photo-interpretation using 50-cm spatial resolution orthophotos and S2 images acquired in 2017.

It should be noted that the photo-interpreted poplar polygons correspond to relatively mature plantations of more than three years old. When the trees are younger, the canopy cover may be partial and poplar stands may be confused with other species. Three experts were responsible for photo-interpretation and a set of poplar polygons from the south-western tile (31TCJ) was validated in the field by forest partners. Among a total of 230 photo-interpreted polygons, only two corresponded to eucalyptus plantations, representing less than 1% misidentification.

3. Methods

The processing chain we proposed is shown in Figure 5.4. It consisted of three main steps: pre-processing, local processing and global processing. For large area mapping, only one local model is required to initiate the learning process. The model is then adapted with AL for spatial transferability, by introducing iteratively a limited number of reference data from other areas. This adaptation is intended to build a global model that can be applied to the new tiles while maintaining good predictive performance on the initial one.

In order to better assess the predictive power of the global model and to measure the impact of choosing the initial model, three local models were also learned, one for each tile. Each of them was used to initiate the global processing with AL, for comparison.

3.1 Pre-processing

The pre-processing step was the same for all the three study tiles. As illustrated in Figure 5.2, the acquisitions are asynchronous between the three tiles as well as the cloud coverage. Therefore, for each Sentinel-2 SITS, all available acquisitions were used and the missing data were filled using a multitemporal linear interpolation (Inglada, 2016). This involves replacing

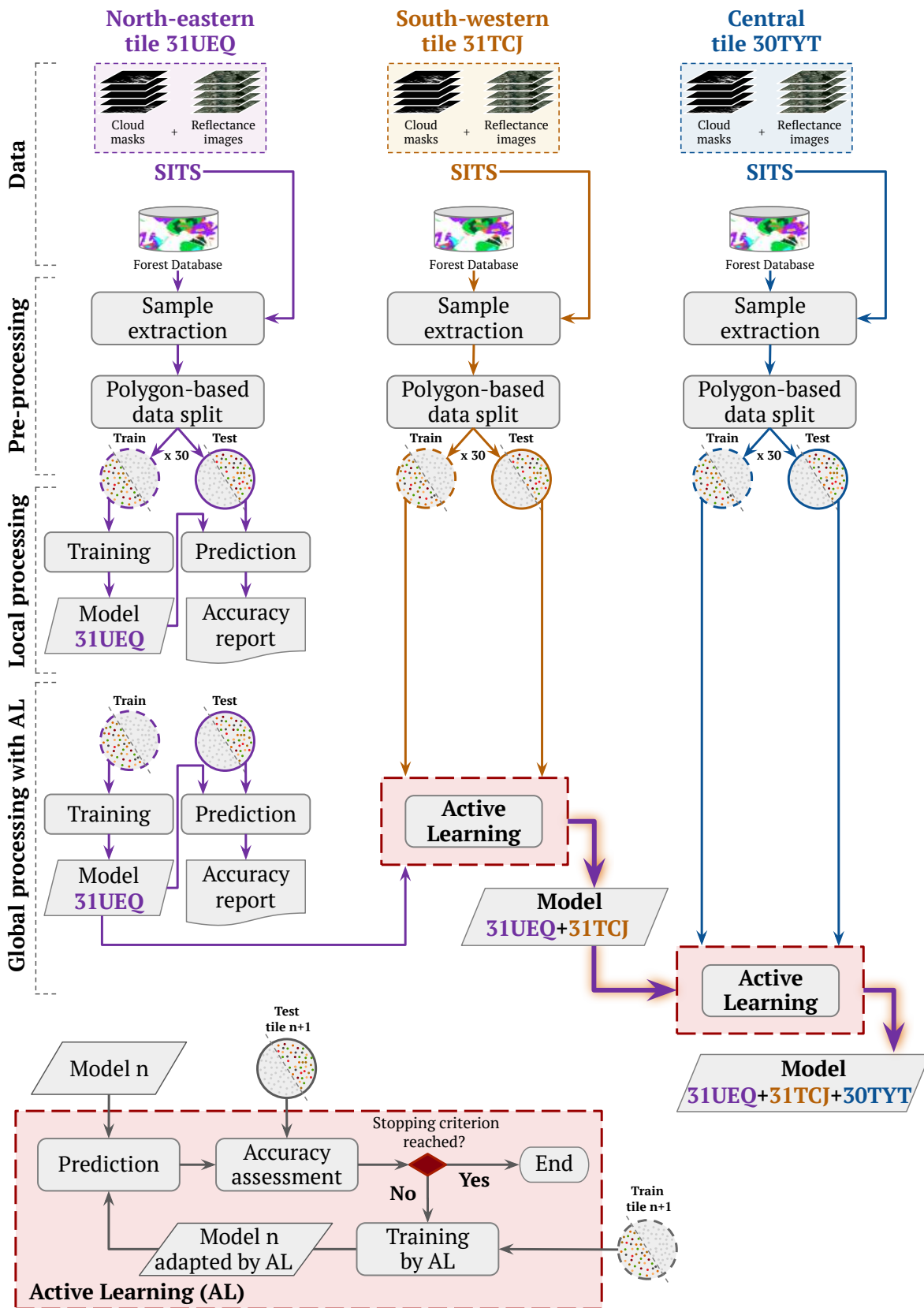


Figure 5.4 – Flow chart of the proposed approach for the classification of large areas by active learning. The chart illustrates an example of a global processing along a transfer direction from the north-eastern (31UEQ) to the south-western tile (31TCJ) and ending with the central one (30TYT).

each invalid pixel (detected in the cloud mask) by an interpolated value with the closest and valid pixels in the SITS (Inglada et al., 2016). The efficiency of the linear gap-filling has already been demonstrated, particularly with a short time gap between image acquisitions (Julien and Sobrino, 2010; Kandasamy et al., 2013). It offers a good compromise between computation time and accuracy (Inglada et al., 2015). An illustration of the gap-filling process is provided in Appendix A.2. The time series were then resampled at 10-m spatial resolution with a 10-day time step common to all tiles. The resulting gap-filled SITS comprise 340 bands corresponding to 10 spectral bands * 34 resampled dates over the course of the year. The precise acquisition dates and the 34 resampled dates are provided in Appendix A.1.

Reference polygons (i.e. forest stands) were used to stratify the pixel sampling. First, the polygons were randomly split, 50% for training and 50% for testing. Then, pixels were sampled in each subset of polygons in the same proportions. Polygon splitting was repeated 30 times to quantify the variability related to the random selection. To reduce spatial auto-correlation (Karasiak et al., 2019), we avoided sampling pixels from the same polygons into training and testing data sets. In order to ensure class balance, training and testing sets of pixels were built using a stratified random sampling of equal size for each class belonging to the same S2 tile. The selection was performed so that the smallest class was fully sampled (except for a few, to have a rounded number). The sample size varied between 2500 and 7700 pixels (i.e. one S2 pixel of 10 m² area) according to the tile. The values of each pixel were extracted from the 340 spectral bands of the gap-filled S2 SITS. Table 5.1 provides an overview of the total number of samples used (for both training and testing) for each class and in each tile.

Table 5.1 – Total number of reference samples per class extracted from each S2 tile. This number is the same for both reference sets (outdated and updated data). The sample set was split to use 50% for training and 50% for testing. Training and testing pixels were collected from spatially disjoint reference polygons (i.e. forest stands).

| Sample size in pixels per class | | | | | | | |
|---------------------------------|--------|--------|-----------------|------|-----------------|---------------------------------|-------------------------------|
| Tile code | Poplar | Locust | Chestnut | Oak | Beech | Closed deciduous forest (mixed) | Open deciduous forest (mixed) |
| 31UEQ | 2500 | 2500 | NA ¹ | 2500 | 2500 | 2500 | 2500 |
| 30TYT | 4000 | 4000 | 4000 | 4000 | NA ¹ | 4000 | 4000 |
| 31TCJ | 7700 | 7700 | 7700 | 7700 | NA ¹ | 7700 | 7700 |

¹ Class not available in the study area or poorly represented.

3.2 Local classification approach

For each of the three S2 processed SITS, a random forest (RF) (Breiman, 2001) supervised classification was carried out independently. The main objective was to assess the ability of

the algorithm to identify poplar plantations at a local scale in the same tile where labelled reference data are sufficiently available. This step allowed also the construction of the best possible local model that would be used to initiate the global learning approach.

The RF classifications were performed using the *scikit-learn* Python library (Pedregosa et al., 2011) with prior optimisation of the algorithm’s hyperparameters: the number of trees, the maximum depth of the trees, and the number of features at each split. The optimisation was conducted using a grid-search approach to find the best combination of the hyperparameters based on a 5-fold cross-validation of training data. The grid search value for the number of trees ranged from 10 to 150 with a step of 10 while the maximum depth varied from 5 to 100 with an interval of 10. The default value for the maximum tree depth was also tested (i.e. *None*). It consisted in expanding the tree nodes to their maximum depth until all the leaves became pure. This value is determined by the algorithm and depends on the number and quality of samples in each node. Finally, the number of features to split a node was fixed as the square root of the total number of input variables (Breiman, 2001; Ghosh and Joshi, 2014; Belgiu and Drăguț, 2016). This value was tested in addition to a range of values from 1 to 20 with a 2-value step. A total of 1694 combinations were thus performed.

Classifications were repeated 30 times corresponding to the 30 random splits of the reference data into 50% for training and 50% for testing. This process was conducted for the three S2 study tiles and the results were assessed for each tile separately. Confusion matrices and performance measures (Overall Accuracy, F-score, Precision and Recall) were computed and averaged over the 30 repetitions. The Kappa (κ) metric was deliberately disregarded (Foody, 2020).

In order to evaluate the importance of checking and cleaning reference samples of poplar plantations by photo-interpretation, the local classification models were similarly built with poplar samples directly derived from the forest database, with no update (outdated data). The same sample size was respected in both data sets (Table 5.1).

3.3 Global classification approach

The previous section detailed the traditional classification scheme where the learning process and the predictive performance are assessed in each tile independently assuming that a relatively large number of reference samples is available for each tile. In this next step, active learning is proposed as a transfer strategy to build a global model from a local one with the addition of a limited number of reference samples from new geographical areas.

3.3.1 Principles of active learning

Active learning is based on the assumption that a machine learning algorithm can perform better with fewer training samples than traditional methods if it is allowed to choose the data from which it learns (Settles, 2012). AL makes iteratively effective queries to label the most informative samples rather than randomly selecting them. These labelled samples are often located in low confidence regions, which helps the initial learner to reduce its uncertainty (Tuia et al., 2011b). The process is iterated until certain predefined stopping criteria are met (Vlachos, 2008), for instance, the maximum number of additional samples or the maximum performance score.

In the AL literature, there are three common scenarios or ways in which the learner may query the labelled samples: query synthesis, stream-based active learning and pool-based active learning. In a *query synthesis* scenario (Angluin, 1988), the learner may either select any unlabelled instance in the input space or generate new instances (from the underlying distribution) and request for their labels. The main problem with this setting is that the new instances generated by the learner may be difficult to label (Settles, 2012). When a *stream-based active learning* scenario (Cohn et al., 1994) is considered, the unlabelled instances are drawn one at a time in a sequential process and the learner decides whether or not to query their labels according to a predefined ranking strategy. The third scenario *pool-based active learning* (Lewis and Gale, 1994) is the most popular in the community (Settles, 2012). It assumes that a large pool of unlabelled data \mathcal{U} is available and only the most useful samples \mathcal{L} are selected to be labelled (Lewis and Gale, 1994; McCallum and Nigam, 1998). The decision whether to query or reject a sample is made according to a ranking strategy that guides the model to choose the best samples (i.e. the most informative for the learning algorithm). The active learner calculates an information metric of all the samples in the entire pool, ranks them according to this value, selects the most informative ones and then, queries their labels (i.e. the classes) to train a new model. Unlike the previous scenario which makes query decisions individually for each instance, in the pool-based setting, all the instances are ranked at the same time and the most informative are then selected.

Two ranking criteria can be considered in AL: *uncertainty* and *diversity*. Sampling using *uncertainty* consists in querying only the instances for which the initial learner is the least certain of their labels after prediction (Lewis and Gale, 1994). Following this strategy, the most informative samples are the most uncertain. For classification tasks, uncertainty is usually quantified by one of three measures: least confident (LC), margin sampling (MS) or entropy (H) (Settles, 2012).

Diversity-based methods tend to exploit the data structure by selecting the samples that are the most representative or diverse of the input space distribution (Settles, 2012). Density-

weighted metrics or clustering-based approaches are generally used and they are frequently combined with an uncertainty metric (Brinker, 2003). Density metrics are based on similarity measures of distance or angle. Samples which minimise similarity values (i.e. the most dissimilar samples) are favoured. Several similarity metrics are used including Euclidean distance (Gong et al., 2019) and Cosine similarity (Zhong et al., 2017).

3.3.2 *Experimental setup for global classification with active learning*

Our global classification scheme with AL starts with an accurate and already trained local model on a first data source (the north-eastern S2 tile in the example shown in Figure 5.4). Then, the initial model (Model 31UEQ) is used to predict the classes in a target domain (e.g. the south-western S2 tile). Without any adaptation, this first external prediction is assumed not to be very robust because of sample selection bias (Persello and Bruzzone, 2012). If the predicted performance based on the initial model does not meet a predefined stopping criterion, new samples from the target domain are queried by AL and added to the initial model to be retrained. This procedure is repeated iteratively until the stopping criterion (here, defined in terms of the number of samples added from the target domain) is reached. The resulting model (Model 31UEQ+31TCJ) is expected to fit both source and target domains well. The same procedure is used in the following step to adapt the two-tile model (Model 31UEQ+31TCJ) to the third tile (30TYT) ultimately leading to a global model tailored to the three S2 tiles (Model 31UEQ+31TCJ+30TYT).

In order to check the symmetrical transferability of the models from one tile to another, we computed all six possible combinations to build the global model (starting from tile 31UEQ or the others). We assumed a pool-based active learning scenario in which new samples from the target training set are queried based on their informativeness. Two uncertainty metrics were tested to select new samples by AL: entropy (H) and margin sampling (MS). These measures can be defined as follows, with x_M^* the best instance selected for a query using an uncertainty measure M from the unlabelled data pool \mathcal{U} :

- *Margin sampling (MS)*: this metric consists in querying the sample with the smallest difference between its two most probable labels (Settles, 2012). The closer the probabilities, the less confident the model (i.e. great doubt between the two labels). The best instance is the one that minimises this value:

$$x_{MS}^* = \operatorname{argmin}_x [P_\theta(\hat{y}_1 | x) - P_\theta(\hat{y}_2 | x)] \quad (5.1)$$

where \hat{y}_1 and \hat{y}_2 are respectively the first and second most probable labels under the model θ and P_θ denotes the probability value with the same model θ .

- *Entropy (H)*: for each sample, the entropy takes the probability of it belonging to all possible model classes into account (Shannon, 1948). When the degree of certainty is high, the entropy value is low (i.e. a high probability of belonging to a specific class). Great uncertainty arises when the probability values are shared between many classes and none stands out (i.e. high entropy). The samples selected by the algorithm are therefore those with the highest entropy value:

$$\begin{aligned}
 x_H^* &= \operatorname{argmax}_x H_\theta(Y | x) \\
 &= \operatorname{argmax}_x - \sum_y P_\theta(y | x) \log P_\theta(y | x)
 \end{aligned}
 \tag{5.2}$$

where H is Shannon’s entropy and y ranges over all possible labels of x .

For the purpose of comparison, diversity metrics were also combined with the entropy and margin sampling uncertainty metrics. Euclidean distance and cosine angle were tested. While Euclidean distance measures the length of the straight line segment connecting two points in Euclidean space, cosine similarity is defined as the cosine of the angle between them. Since AL with the combined uncertainty and diversity metrics performed worse, these results are listed in Appendix A.6.

In all the experiments, we considered only samples from the pure-species classes. Mixed classes were removed from the target training set assuming that they could negatively affect the AL process (as reported in Section 4.2.5 and discussed in Section 5.5). AL was run by adding 10 target samples at each iteration up to a maximum of 1000. From an operational point of view, this maximum number is excessive but in these experiments, it enabled us to understand the impact of increasing the size of the training set in AL.

In order to assess the contribution of AL, target samples were also queried using a random sampling (RS) strategy with the same sample size at each iteration. For both classification procedures (AL *vs.* RS), the performance metrics were calculated on testing sets of both source and target domains and averaged over the 30 repetitions. The entire AL workflow was implemented in Python using the *modAL* framework (Danka and Horvath, 2018).

The F-score per class (i.e the harmonic mean of precision and recall) was selected to conduct the experiments and present the results. However, since this metric is affected by class prevalence, a summary of the results is also provided in Appendix A.5 with the Recall (or producer’s accuracy) and Precision (or user’s accuracy) metrics, leading to the same conclusions. An imbalance between classes appears when the poplar class is compared to all the others (i.e. one against all configuration).

4. Results

4.1 Tile by tile classification: local approach

Table 5.2 summarises the results of the random forest classification for each of the three study tiles using the 2017 Sentinel-2 SITS. As a reminder, these best local models are intended to initiate the global active learning-based approach. They also inform on the ability to discriminate poplar plantations from the other deciduous species when enough labelled reference data are available.

Table 5.2 – Results of site-specific classifications for each S2 tile averaged over 30 independent repetitions.

| Tile code | Training size ¹ per class in pixels | No. classes | Overall Accuracy _(*30) | Global F-score _(*30) | Poplar F-score _(*30) | Poplar Precision _(*30) | Poplar Recall _(*30) |
|--|---|-------------|-----------------------------------|---------------------------------|---------------------------------|-----------------------------------|--------------------------------|
| Without photo-interpretation of poplars (outdated data) | | | | | | | |
| 31UEQ | 1250 | 6 | 65.6±6.5 % | 65.2±6.9 % | 72.6±5.7 % | 69.5±8.2 % | 77.3±3.9 % |
| 30TYT | 2000 | 6 | 65.8±2.2 % | 70.6±2.2 % | 86.7±1.7 % | 85.6±2.8 % | 87.9±2.4 % |
| 31TCJ | 3850 | 6 | 79.5±3.3 % | 79.4±3.7 % | 89.1±3.9 % | 88.6±5.2 % | 90.8±3.2 % |
| With photo-interpretation of poplars (updated data) | | | | | | | |
| 31UEQ | 1250 | 6 | 73.7±1.7 % | 73.1±2.0 % | 89.5±3.3 % | 90.2±6.7 % | 89.2±3.2 % |
| 30TYT | 2000 | 6 | 74.9±1.8 % | 75.0±1.9 % | 99.3±0.2 % | 99.4±0.5 % | 99.3±0.6 % |
| 31TCJ | 3850 | 6 | 80.0±0.7 % | 80.1±0.6 % | 97.9±0.8 % | 99.3±0.5 % | 96.5±1.7 % |

¹ Training samples represent 50% of the available reference data.

When all reference polygons were retrieved from the forest database, including poplar samples without being photo-interpreted, an average loss of about 17%, 13% and 9% of the poplar F-score was observed for the north-eastern, central and south-western tiles respectively. After checking and updating poplar samples by photo-interpretation, the results revealed a high capacity of identifying poplar plantations at the tile scale with an average F-score ranging from 89.5% to 99.3%.

Considering the other classes, accuracies were lower but slightly better after photo-interpretation of poplars, leading to average values of global F-score (including all the classes) between 73.1% and 80.1%, depending on the S2 tile considered.

4.2 Global classification approach

4.2.1 Active learning vs. random sampling: overall accuracy assessment

As mentioned above, two active learning uncertainty measures: entropy (AL_H) and margin sampling (AL_{MS}) were compared to a random selection of samples (RS). For a transfer between two tiles, six models corresponding to the six possible combinations resulted from the three study tiles. The average overall accuracy (OA) values of these two-tile models are reported in Table 5.3 according to the transfer direction, the number of added samples and the approach adopted (RS, AL_H and AL_{MS}). In Table 5.4, we present the OA values achieved

with three-tile-based global models for active learning (AL_{MS}). These results are based on the six two-tile models derived from Table 5.3.

Table 5.3 – Average overall accuracy values (in %) according to the transfer direction from one tile to another and the number of target samples added to the initial set for active learning with entropy (AL_H) and margin sampling (AL_{MS}) uncertainty metrics. Comparisons were made with an adapted model including additional randomly selected (RS) target samples. The values are averaged over 30 independent repetitions.

| | | Target tiles | | | | | | | | | | | | | | |
|--------------|------------------------|------------------------------|-----|-----|-----|------|-------|-----|-----|-----|------|-------|-----|-----|-----|------|
| | | 31UEQ | | | | | 30TYT | | | | | 31TCJ | | | | |
| Source tiles | OA (%) | Number of additional samples | | | | | | | | | | | | | | |
| | | 0 | 250 | 500 | 750 | 1000 | 0 | 250 | 500 | 750 | 1000 | 0 | 250 | 500 | 750 | 1000 |
| 31UEQ | <i>RS</i> | - | - | - | - | - | 36 | 50 | 57 | 60 | 61 | 47 | 63 | 69 | 71 | 72 |
| | <i>AL_H</i> | - | - | - | - | - | 36 | 50 | 56 | 60 | 62 | 47 | 59 | 65 | 69 | 71 |
| | <i>AL_{MS}</i> | - | - | - | - | - | 36 | 54 | 61 | 63 | 64 | 47 | 68 | 73 | 74 | 75 |
| 30TYT | <i>RS</i> | 31 | 45 | 52 | 55 | 57 | - | - | - | - | - | 53 | 69 | 70 | 71 | 72 |
| | <i>AL_H</i> | 31 | 44 | 50 | 53 | 56 | - | - | - | - | - | 53 | 67 | 70 | 71 | 72 |
| | <i>AL_{MS}</i> | 31 | 49 | 55 | 58 | 59 | - | - | - | - | - | 53 | 70 | 72 | 73 | 74 |
| 31TCJ | <i>RS</i> | 40 | 50 | 58 | 61 | 63 | 52 | 61 | 64 | 64 | 65 | - | - | - | - | - |
| | <i>AL_H</i> | 40 | 50 | 59 | 63 | 63 | 52 | 62 | 63 | 64 | 65 | - | - | - | - | - |
| | <i>AL_{MS}</i> | 40 | 53 | 61 | 64 | 65 | 52 | 63 | 65 | 66 | 66 | - | - | - | - | - |

In all cases, a model trained on a source tile and applied to a second target tile showed low OA values before adding new samples. The performances improved as the model was enriched with samples from the target.

Generally, for AL, we observed a better performance with the margin sampling uncertainty metric (AL_{MS}) than with the entropy metric (AL_H) with an OA difference ranging from +1% to +9%. We also observed that the classification performance was better with AL, especially AL_{MS} , than with random sampling (RS). Additionally, performances varied according to the transfer direction and the results showed asymmetrical transferability, particularly when only two tiles were used to build the global model. For instance, an average OA score of 61% was obtained with AL_{MS} from 31UEQ to 30TYT by adding an extra 500 samples versus 55% in the opposite case (Table 5.3).

An illustration of how OA scores fluctuated as a function of the additional target samples is given in Figure 5.5, in the particular case of a transfer from the north-eastern tile (31UEQ) to the south-western tile (31TCJ). At first (i.e. with no additional samples from the target), the original 31UEQ model provided predictions on the 31TCJ tile with an average OA of 47% (*vs.* 80% when a local model was used for the target tile, see Table 5.2, Section 4.1). In that case, the OA scores increased as the target samples were added randomly or by AL, with an advantage for AL_{MS} , especially with few extra samples. On average, AL_{MS} significantly outperformed the RS model by 3.5% (Wilcoxon signed-rank test with a p -value < 0.01). A maximum difference of 5.5% between the two learning curves (AL_{MS} *vs.* RS) was also

found based on the Hausdorff distance. Concerning entropy-based active learning (AL_H), the approach yielded results that closely resembled those of RS.

As shown by the dotted lines, the adaptation of the models for the target domain (tile 31TCJ) does not reduce the classification accuracy in the source domain (tile 31UEQ). The OA scores remained fairly constant when target samples were added (Figure 5.5). The same trend was observed regardless of the learning approach (RS or AL).

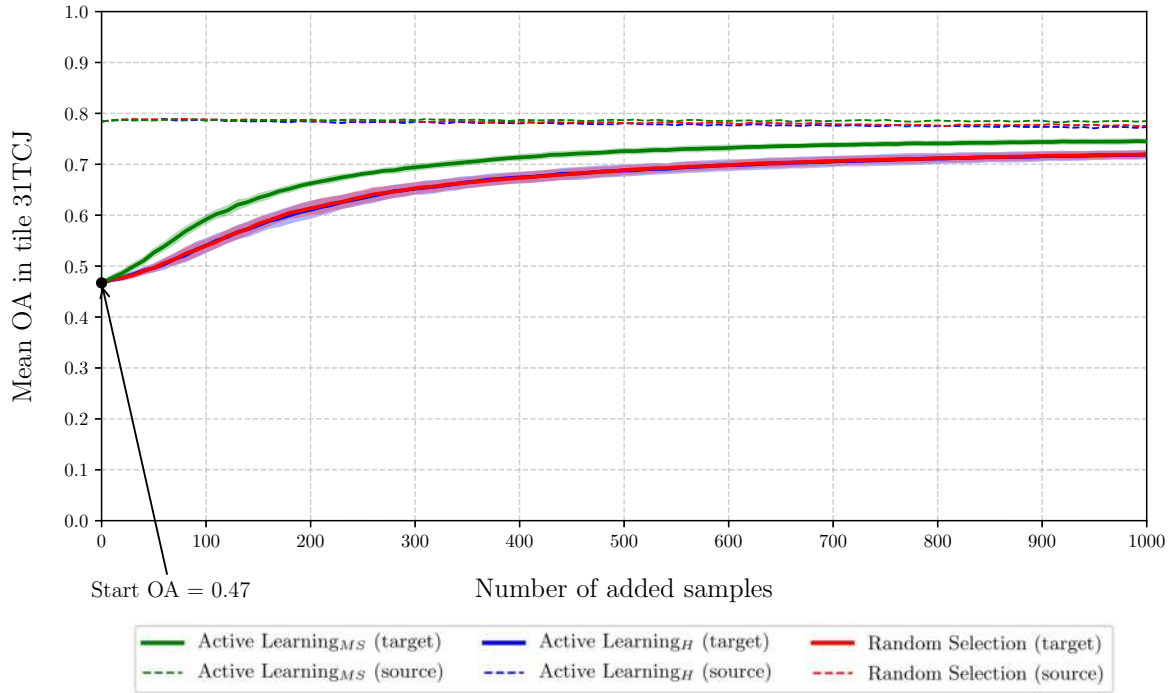


Figure 5.5 – Changes in the average OA scores (\pm standard deviation) on the south-western target tile (31TCJ) according to the additional number of training samples from the initial model (north-eastern source tile 31UEQ). The average values were based on 30 independent runs for both random sampling (RS in red) and active learning using entropy (AL_H in blue) or margin sampling (AL_{MS} in green). Dashed lines show changes in the OA on the north-eastern source tile (31UEQ) with RS (red dashed line), AL_H (blue dashed line) and AL_{MS} (green dashed line).

When the global models included samples from the three S2 tiles (i.e. three-tile models) and considering the approach with the best two-tile results (AL_{MS}), the classification performances were closer (Table 5.4). This is particularly true when the north-eastern tile (31UEQ) was included in the two-tile-based models used to predict the third tile. From 500 extra samples, we found a difference in OA varying by only 1% or 2% between target tiles 30TYT and 31TCJ. For example, an OA of 88% was achieved with 750 extra samples for 30TYT versus 87% with the same number of additional samples for 31TCJ. By contrast, when the north-eastern tile 31UEQ was the target, the three-tile-based models were less accurate with a difference in OA of around 15% compared to the target tiles 30TYT and 31TCJ. Nevertheless, for all three tiles, the maximum achievable performance appeared to be reached with three-tile-based models and exceeded that of local models by 1% up to 14% depending on the tile considered (see

Table 5.2 for comparison). Moreover, the asymmetry between the two source models used to predict each of the 30TYT or 31TCJ target tiles was very small after sample addition (OA differences from 0% to 2%) whereas it was greater when the 31UEQ tile was the target (OA differences from 3% to 10%). For all three target tiles, these differences between each two source models tended to converge towards 1000 samples.

Table 5.4 – Average overall accuracy values (in %) according to the transfer direction from a two-tiles based model (i.e. an initial local model adapted by AL with 1,000 extra samples from a second tile) to a third one. Target samples were added to the initial set by active learning based on margin sampling (AL_{MS}). The values are averaged over 30 independent repetitions.

| Source tiles | OA (%) | Target tiles | | | | | | | | | | | | | | |
|-----------------|-----------|------------------------------|-----|-----|-----|------|-------|-----|-----|-----|------|-------|-----|-----|-----|------|
| | | 31UEQ | | | | | 30TYT | | | | | 31TCJ | | | | |
| | | Number of additional samples | | | | | | | | | | | | | | |
| | | 0 | 250 | 500 | 750 | 1000 | 0 | 250 | 500 | 750 | 1000 | 0 | 250 | 500 | 750 | 1000 |
| 30TYT +31TCJ | AL_{MS} | 36 | 61 | 71 | 76 | 78 | - | - | - | - | - | - | - | - | - | - |
| 31TCJ +30TYT | AL_{MS} | 33 | 51 | 65 | 71 | 75 | - | - | - | - | - | - | - | - | - | - |
| 31UEQ +31TCJ | AL_{MS} | - | - | - | - | - | 50 | 79 | 85 | 88 | 89 | - | - | - | - | - |
| 31TCJ +31UEQ | AL_{MS} | - | - | - | - | - | 57 | 79 | 85 | 87 | 88 | - | - | - | - | - |
| 31UEQ +30TYT | AL_{MS} | - | - | - | - | - | - | - | - | - | - | 47 | 82 | 87 | 89 | 90 |
| 30TYT +31UEQ | AL_{MS} | - | - | - | - | - | - | - | - | - | - | 52 | 84 | 87 | 89 | 90 |

4.2.2 Active learning vs. random sampling: class-specific assessment

In order to better understand what underlies the OA measurements, the F-score values by class were analysed in relation to the total number of additional samples (see Appendix A.5 for non-prevalence-dependent Recall metric and for Precision metric). For the sake of simplicity and clarity, results are reported here for only one transfer direction.

Considering a transfer from the north-eastern tile (31UEQ) to the south-western tile (31TCJ) with no extra training samples, the F-score obtained for poplar plantations was 92% (Figures 5.6a – 5.6c). This score increased with the addition of new samples to reach a maximum at 97% and 98% respectively, for RS and both active learning models (AL_{MS} and AL_H).

Regarding the class of locust tree species, a low F-score of 7% was noted prior to the addition of target training samples. However, this value rose sharply to achieve 59%, 32% and 67% with 250 extra samples for respectively RS, AL_H and AL_{MS} models (Figures 5.6a – 5.6c). In the same way, the initial F-score for chestnut tree species was null in the target

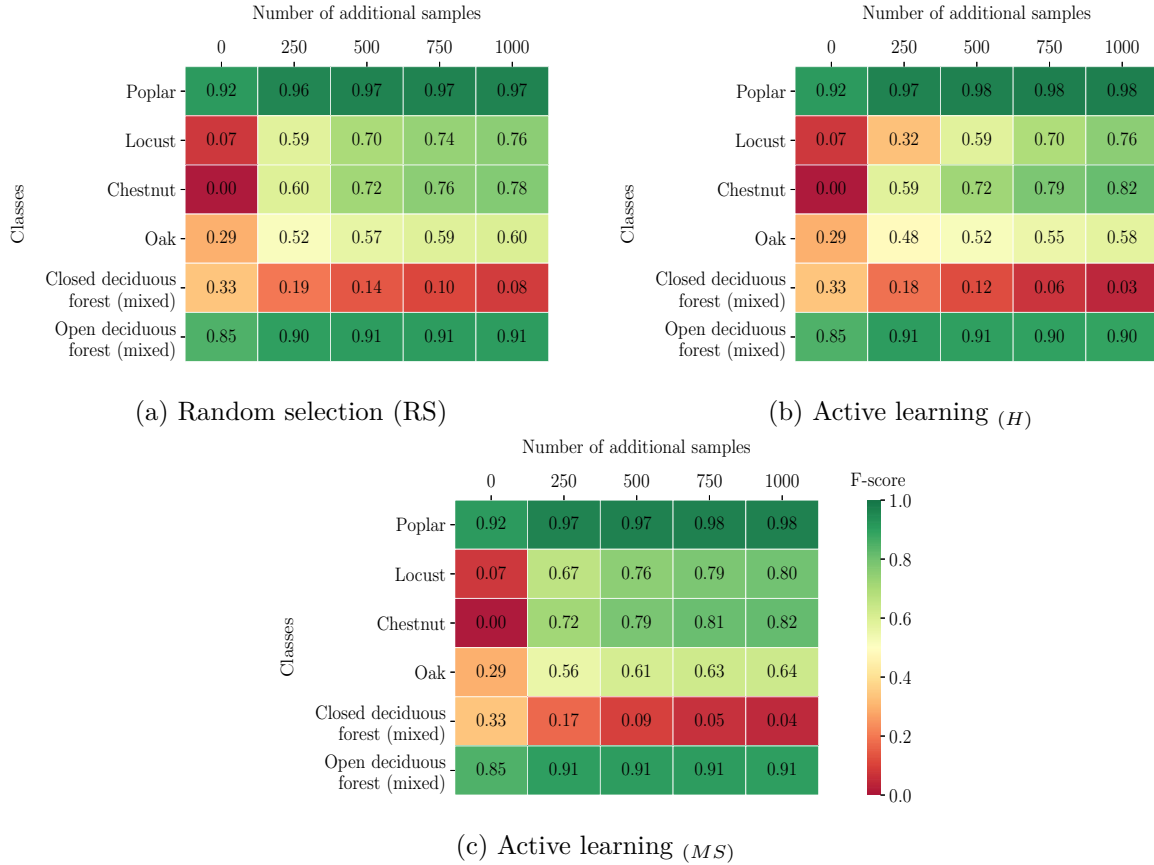


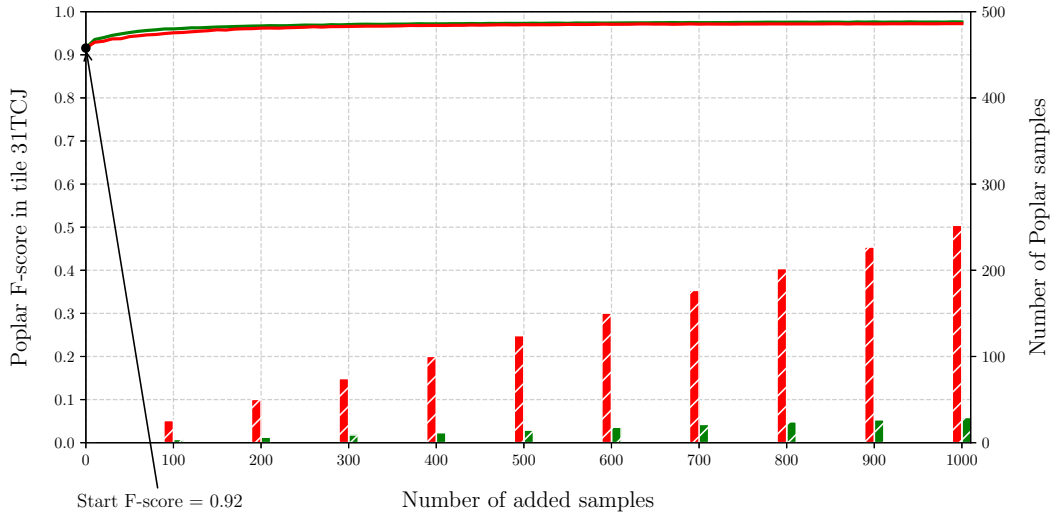
Figure 5.6 – Average F-score values per class obtained over 30 independent repetitions as a function of the number of added samples with (a) random sampling, (b) entropy-based active learning (AL_H) and (c) margin sampling-based active learning (AL_{MS}). The transfer was performed from the north-eastern tile (31UEQ) to the south-western tile (31TCJ).

domain because of the absence of this class in the source domain. However, after a few iterations, the accuracy jumped to around 60% for both RS and AL_H models (250 extra samples). Like in the previous case, the rate of increase with the AL_{MS} model was faster and reached an F-score value of 72% at the 25th iteration. In these last two examples (locust and chestnut tree species), RS required up to 50% more samples to achieve the same values of class F-score as AL_{MS} . The same behaviour was observed for oak and open deciduous forest classes with an upward trend in the F-score when training samples were added. The improvement was particularly pronounced when the initial value was low. This was true except for the closed deciduous forest class (with mixed species) for which the F-score dropped with the addition of new samples whatever the learning approach (RS *vs.* AL) used.

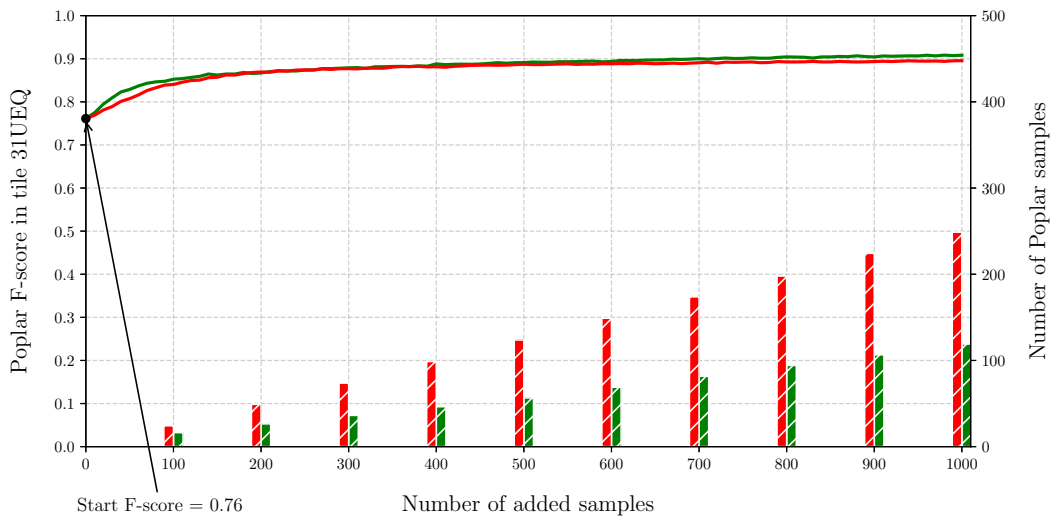
4.2.3 Impact of the transfer direction: case of the poplar class

When the transfer was conducted from the north-eastern (31UEQ) to the south-western (31TCJ) tile, the poplar F-score was equal to 92% before adding any extra samples from the target (Figure 5.7a). This value increased slightly to level off at 97% after the addition of

75 and 8 new poplar samples out of 300 additional samples for respectively RS (red bars) and AL_{MS} (green bars). Although the F-score reached its maximum value from the 20th iteration, the random model continued to select poplar samples to achieve 250 extra samples out of 1000 compared to 26 with the AL_{MS} model.



(a) From 31UEQ to 31TCJ



(b) From 31TCJ to 31UEQ



Figure 5.7 – Changes in the average poplar F-score values between the north-eastern (31UEQ) and south-western (31TCJ) tiles according to the number of added samples. In (a), the change in the average poplar F-score is given on the south-western target tile (31TCJ) with an initial classifier trained on the north-eastern source tile (31UEQ). In (b), the change in the average poplar F-score is given on the north-eastern tile based on the south-western model. Red (random sampling) and green (sampling based on active learning with the MS uncertainty metric) bars show the number of poplar samples selected from the target (right y-axis) out of a total of 1,000 samples queried.

When the transfer across space was performed in the opposite direction (i.e. from the south-western 31TCJ tile to the north-eastern 31UEQ tile), the result was slightly different.

With no additional samples from the target, the poplar F-score was 76% (Figure 5.7b). It increased to achieve a stable value around 90% with 50 and 100 extra poplar samples (out of 400 additional samples) for AL_{MS} and RS models, respectively. In the first iterations, AL_{MS} queried more poplar samples than in the opposite direction. However, the number was still lower than with a random selection of samples (RS).

4.2.4 Impact of missing classes: case of the chestnut class

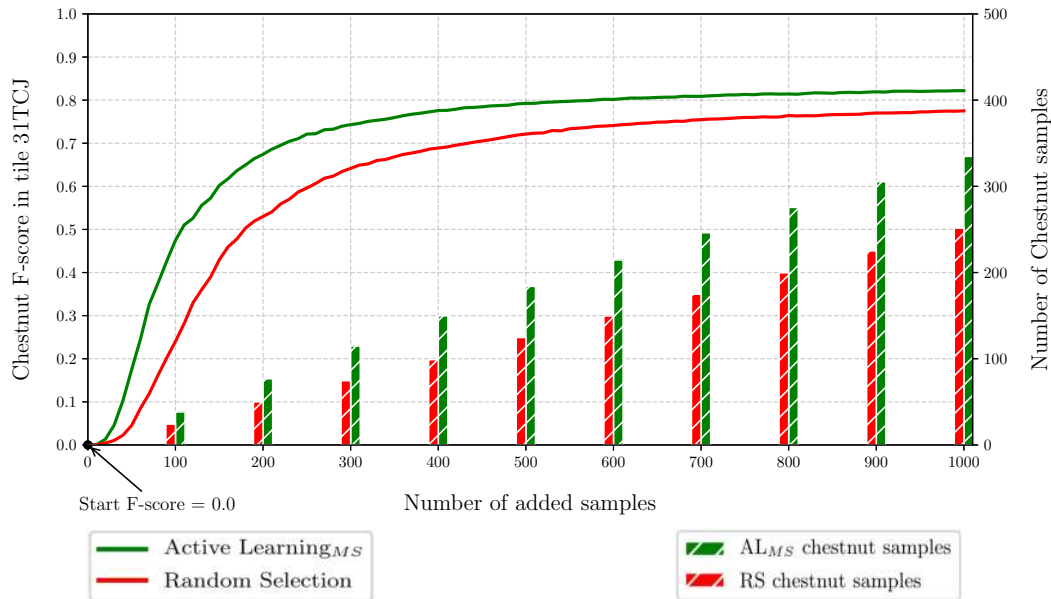


Figure 5.8 – Changes in the average chestnut F-score values on the south-western target tile (31TCJ) according to the number of added samples. The initial classifier was trained on the north-eastern source tile (31UEQ).

In the case of a transfer from the north-east to the south-west, the original model does not include the chestnut class which actually exists in the target tile (Table 5.1). As illustrated in Figure 5.8, the initial chestnut F-score was equal to zero. This value increased rapidly with the addition of extra chestnut samples from the target for both RS and AL_{MS} models but considerably faster with AL_{MS} which reached up to 20% higher F-score. The number of additional chestnut samples actively selected was also greater than that added with RS.

4.2.5 Impact of including mixed classes in the training set: case of the locust class

As mentioned in Section 3.3.2, all the models were trained using only pure pixels of tree species. To evaluate the impact of considering mixed classes, we compared the performance of the random model (RS) and the two active learning-based models (AL_H and AL_{MS}) with and without including mixed classes in the training sets. Results are presented in Table 5.5

Table 5.5 – Average locust F-score on the south-western tile according to the total number of additional samples with and without the presence of mixed classes in the training set.

| Average locust F-score (%) _(*30) | Number of additional samples (total) | | | | | | | | | |
|--|--------------------------------------|-----|-----|-----|------|-----------------------|-----|-----|-----|------|
| | With mixed classes | | | | | Without mixed classes | | | | |
| | 0 | 250 | 500 | 750 | 1000 | 0 | 250 | 500 | 750 | 1000 |
| Random Selection | 7 | 40 | 59 | 67 | 71 | 7 | 59 | 70 | 74 | 76 |
| Active Learning (<i>H</i>) | 7 | 12 | 25 | 34 | 43 | 7 | 32 | 59 | 70 | 76 |
| Active Learning (<i>MS</i>) | 7 | 52 | 65 | 70 | 73 | 7 | 67 | 76 | 79 | 80 |

for the locust species class following a transfer from the north-eastern source tile (31UEQ) to the south-western target tile (31TCJ).

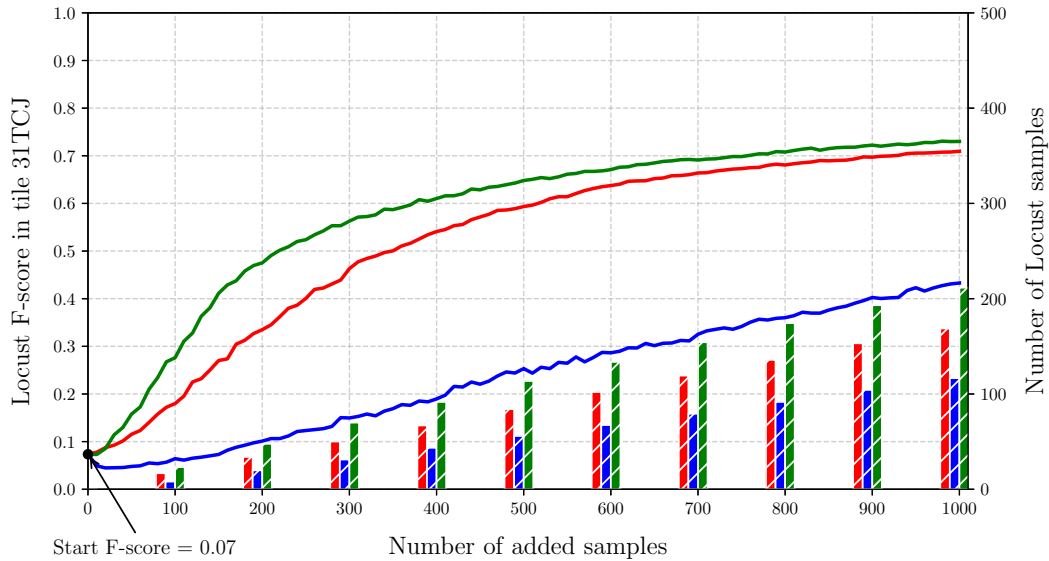
Before adding extra samples from the target, the locust F-score was equal to 7%. When mixed classes were included in the training set, the locust F-score improved as new samples were added especially for AL_{MS} and RS to achieve respectively 73% and 71% at the last iteration (with a total of 1000 extra samples). The lowest performance was achieved with AL_H and the number of locust samples selected from the target was also higher for AL_{MS} and RS than for AL_H (Figure 5.9a).

When mixed classes were excluded from the training set, the performance improved faster than in the previous case and reached better locust F-score values of 76% and 80% for both AL_H and RS and for AL_{MS} respectively. Although in the latest iterations AL_H has caught up with RS, in the first selections the results of RS outperformed those of AL_H . For instance, at the 40th iteration, the locust F-score was 51% for AL_H versus 68% for RS with respectively 80 and 99 locust samples selected (Figure 5.9b).

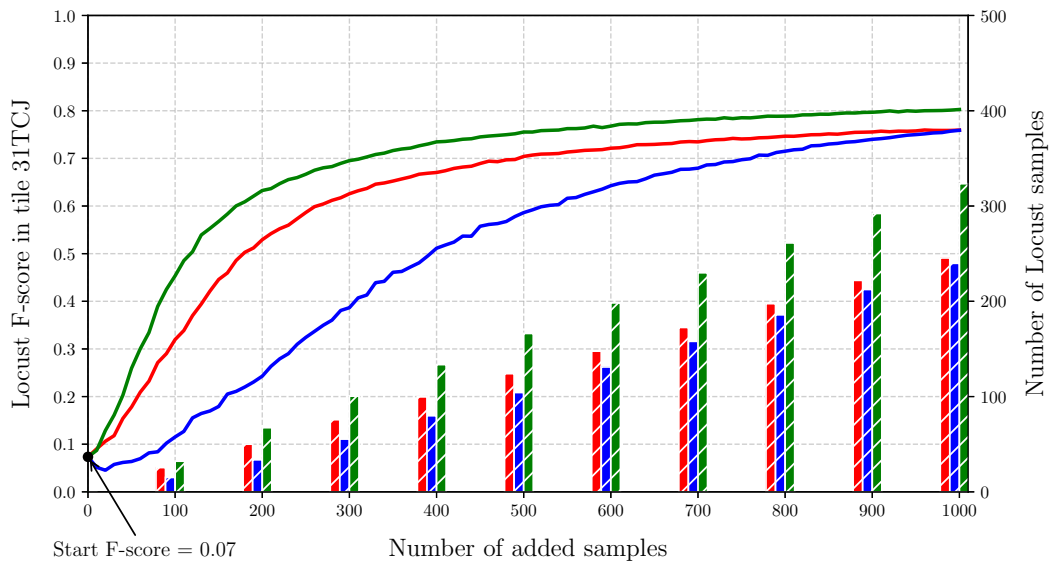
For the three models, the number of locust samples selected from the target has also increased compared to the setting involving mixed classes.

4.3 Maps of poplar plantations

The three S2 tiles were predicted with non-adapted models (before AL) and with a three-tile global model offering the best performances (31UEQ+31TCJ+30TYT). The results are shown in Figure 5.10 with extracts from the three study tiles. Consistent with the statistical metrics, cartographic results are improved using the global model. Confusions are reduced with less under and over-detection of poplars. Complete prediction maps of the three tiles are provided in Appendix A.7.



(a) Training with mixed classes



(b) Training without mixed classes



Figure 5.9 – Changes in the average F-score values on the south-western target tile (31TCJ) for the locust class according to the number of additional samples. The initial classifier was trained on the north-eastern source tile (31UEQ). In (a) the training set contained pure and mixed species while in (b) only pure classes were considered.

5. Discussion

5.1 Importance of up-to-date references

Our results highlighted the importance of having updated reference data for learning instead of sampling them directly from the existing digital forest maps, with no cleaning step. Without checking poplar samples by photo-interpretation, the average F-score for poplars

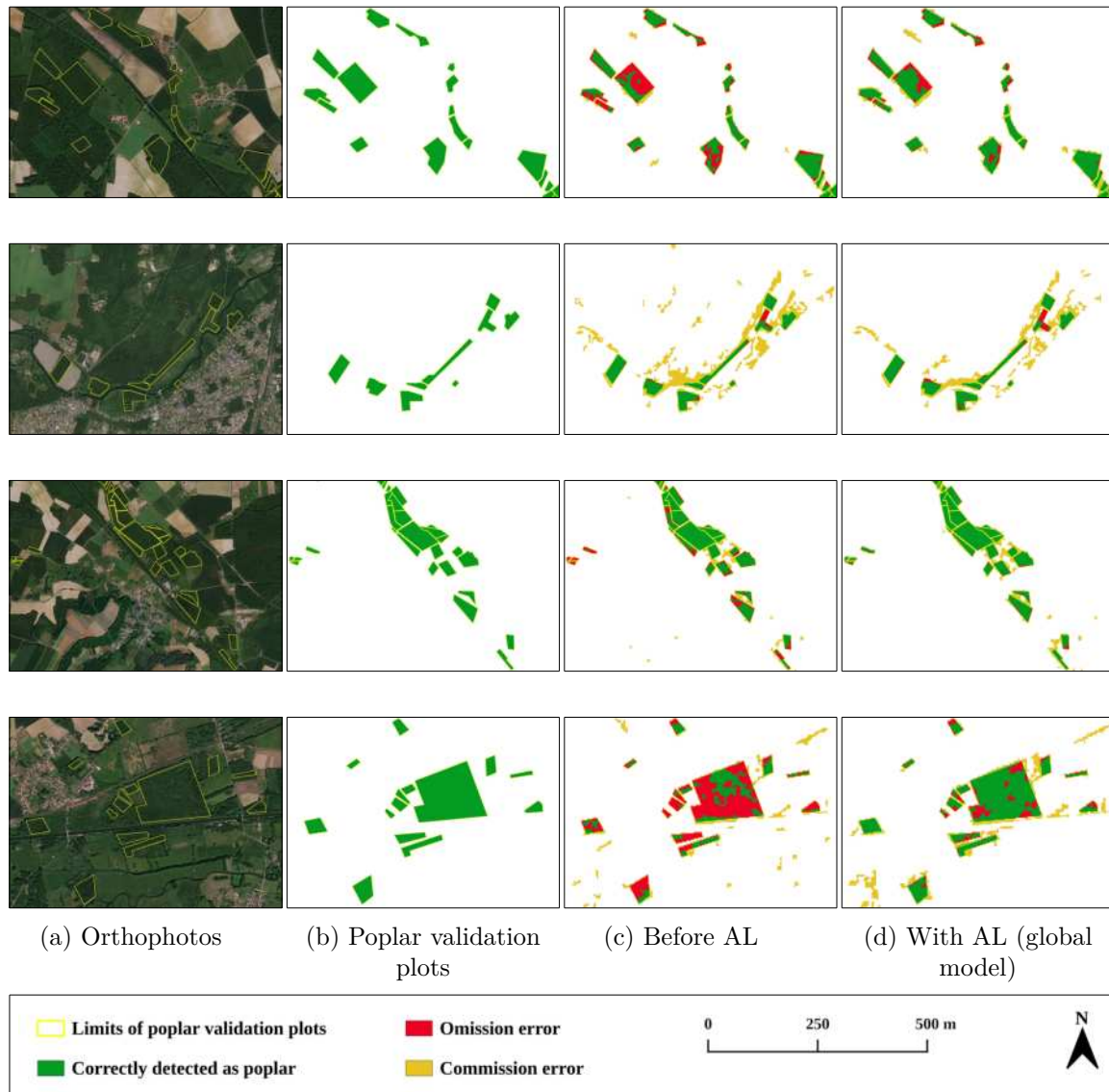


Figure 5.10 – Four map extracts zoomed on poplar plantations comparing the predictions achieved before and after adaptation by AL. (a) Orthophotos of each zoomed area. (b) Poplar validation plots. (c) Classification results according to source tile models before adaptation with AL. (d) Classifications predicted with the three-tile global model built by AL.

dropped from 9% to 17% according to the tile considered. The impact was more noticeable on the north-eastern tile and could be related to the older acquisition years of the aerial photographs used to generate the forest maps (from 2004 to 2013). The poplars derived from these maps presented various errors corresponding to logged plantations or grasslands. The results were therefore statistically and cartographically unsatisfactory and they did not accurately reflect the reality on the ground.

Although the photo-interpretation step was beneficial for the poplar class, the overall improvement in the quality was reduced. This result was unsurprisingly expected as the photo-interpretation was not performed for the other tree species. The difference in OA values between the three tiles can similarly be explained by the difference in the acquisition

years of the forest maps orthophotos with an advantage for the south-western tile. It can also be related to the variation between the classes of tree species and their composition within the three tiles and therefore the classification difficulties are not specifically the same. This discrepancy is slightly reduced with updated references.

Because human expertise is required for checking forest stands from the national forest map, it underlines the interest of adopting an active learning strategy to map poplar plantations over large areas.

5.2 Active learning is an efficient way to generalise across space with limited extra target samples

Our results also revealed low classification performances with no adaptation of the initial local models due to the non-stationarity of class distributions between the different study tiles. After querying samples from the target domain, the models improved their generalisation capabilities across space. This was true for both random sampling and active learning. However, compared to RS, the number of added target samples was much lower with AL (and consequently, the cost of sample acquisition) while reaching the same level of accuracy. In some cases, the number of extra samples was twofold higher with RS. For instance, to reach an average chestnut F-score of 72%, AL_{MS} required 250 samples versus 500 with RS.

In addition, when the same number of extra samples was added, AL_{MS} outperformed RS from 1% to 12% of the F-score with respect to the transfer direction and the class considered. These results which are in line with those of [Alajlan et al. \(2014\)](#), show that AL is a good way to minimise the number of samples for domain adaptation while maintaining high discrimination capabilities. The sample selection bias is corrected and the global model, which is adapted for large area mapping, is able to match the distributions of the local tiles in both training and test sets. However, to benefit from AL, the uncertainty measure must not be arbitrarily chosen. AL_H produced limited performances, with results that were very close to those obtained with RS. The entropy metric is highly dependent on unlikely classes making it less robust, as reported previously ([Rajan et al., 2008](#); [Joshi et al., 2009](#)). Although entropy is an intuitive indicator of uncertainty, the margin sampling (MS) metric is a more reliable and direct way to estimate uncertainty among the most confused classes ([Joshi et al., 2009](#)).

Furthermore, when we compared the performances of the global models to those achieved with tile-specific local models, the OA values improved by up to +14%. Thus, with this cascade adaptation by AL, we have not only built a model suitable for three different tiles, but we have also exceeded the performance reached with the local models trained with all the samples available in each tile.

5.3 Transfer asymmetry tends to disappear with the global model

Our results showed asymmetrical transferability of the models from one tile to another. This means that the classification accuracy may vary according to the tile used to initiate the learning process. When we considered a transfer from the north-eastern (31UEQ) to the south-western (31TCJ) tile, a high F-score value for poplar plantations was achieved without any additional target samples. The original model was already able to correctly identify (at 92% F-score) south-western poplars with the knowledge acquired from the north-eastern tile (i.e. no adaptation was required). Moreover, the improvement in the F-score was almost the same between AL_{MS} and the random sampling-based model after the addition of new samples. By contrast, the number of poplar samples selected over the total number of additional samples revealed a marked difference. For an equivalent F-score value, the number of samples randomly selected (see red bars in Figure 5.7a) was about eight times higher than those selected by AL_{MS} (see green bars in Figure 5.7a) indicating that extra training samples are highly redundant with random sampling. Here, AL minimises the need to hand-label target poplar samples without sacrificing the classification performance. As highlighted previously, when the class accuracy is high, the active learner avoids querying irrelevant samples (Di and Crawford, 2011).

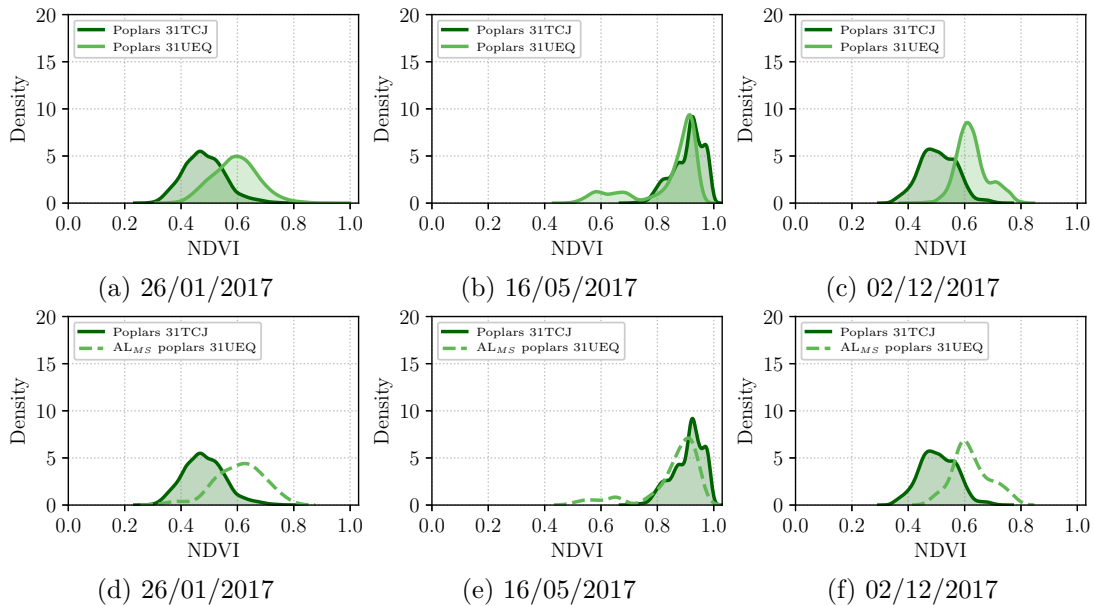


Figure 5.11 – Kernel density of NDVI for poplars in the south-western tile (31TCJ, dark green solid line) and the north-eastern tile (31UEQ, light green solid line) for January 26th in (a) and (d), May 16th in (b) and (e) and December 2nd in (c) and (f), in 2017. In (d), (e) and (f) we added the NDVI distribution of the first 50 poplar samples selected by AL_{MS} (light green dashed line) from the north-eastern target tile when the transfer was conducted from the south-western tile.

In the opposite transfer direction, the south-western model struggled a little with the classification of north-eastern poplars with an initial F-score of 76%. Plantations were mainly

confused with closed deciduous forests (see confusion matrix in Figure A.4a). In this direction, the local environmental conditions of the north-eastern tile may have affected the transferability of the original model and in particular, soil conditions together with related silvicultural practices. Indeed, among the essential requirements for the successful development of poplar plantations, is a well-drained soil with easy access to water (Duchauffour, 1955; Fischer et al., 2018). To satisfy this condition, tillage practices are applied in the driest regions (i.e. south tiles) to minimise competition for water from other plants (shrubs and herbaceous species) resulting in a limited ground cover by the understory vegetation. In the north of France, water is not a limiting factor and most poplar plantations present a grassy or even shrubby cover. Thus, the predictors cover a wider range of variability in the north-eastern tile which makes the classification less accurate without a minimal adaptation, due to the sample selection bias leading to different but related source and target domains between tiles. Stands with dense understory vegetation were almost invisible in the original south-western model and were therefore confused with the closed deciduous forest class in the north-eastern tile. This was confirmed by examining the nature of samples added in the first iterations of AL.

As illustrated in Figure 5.11, the NDVI distribution of the first 50 poplar samples added by AL_{MS} coincides with the uncertainty regions of the south-western model, when the transfer was carried out from the south-west to the north-east.

The asymmetry of the transfer direction began to stabilise with three tiles especially when the north-eastern tile was no longer the target (Table 5.4). This could be attributed to the fact that all the classes are represented in the source tiles (i.e. with the north-east and south-west or the north-east and centre models) unlike the case when the north-eastern tile is predicted (the beech class does not exist in the original models).

Hence, it might be better to start with a two-tile model including the north-eastern tile and then predict either the south-western or central tile in order to ensure the best classification performance in all three tiles. Following the three-tile model, all the class diversity and distribution should already be accounted for (the sample selection bias should disappear), the stability level should increase and equivalent performances would be expected regardless of the order of the tiles.

5.4 Active learning focuses on the most uncertain classes

When the north-eastern source model was directly applied to predict the south-western tile and before adding samples from the target, the chestnut F-score was not surprisingly equal to zero since the chestnut class was initially unknown. All its samples were confused with the pre-existing classes but mainly with closed deciduous forest (51%) and oak (40%) (Figure A.2b). Both AL_{MS} and random models drew chestnut samples from the entire selected pool over the

iterations but many more with the former. In fact, chestnut samples in AL_{MS} were considered to be the most uncertain and were therefore and primarily the most frequently chosen. As discussed in Crawford et al. (2013), the active learning followed an intuitive approach and immediately moved towards this unknown class supposedly difficult to classify. In either case, the F-score increased as chestnut samples were added from the target but considerably faster with AL_{MS} which promptly improved the class definition in the initial source model that successfully identified it (Figure 5.8).

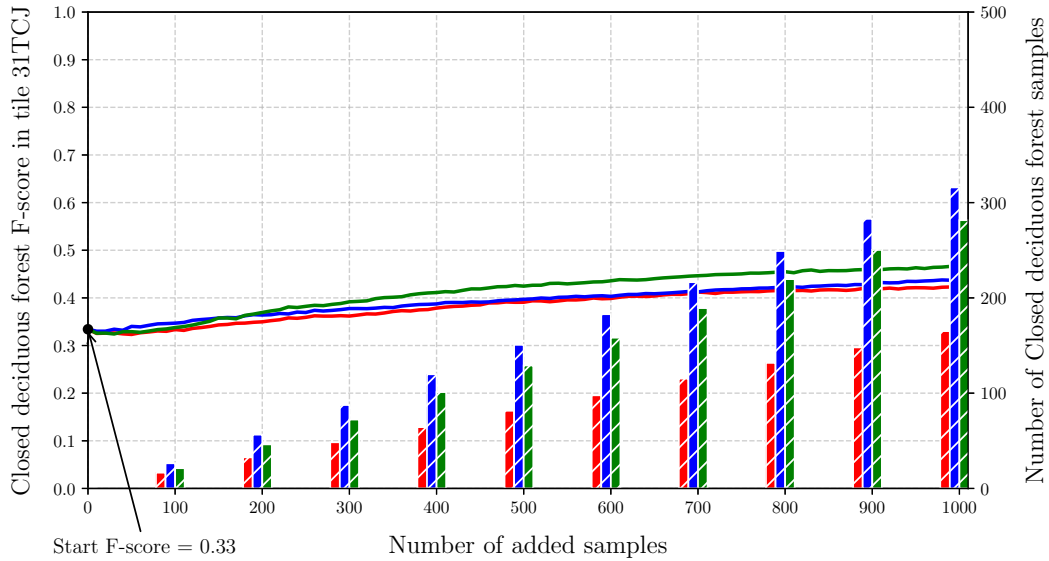
This behaviour of AL was also observed for other missing classes with respect to the transfer direction. These results confirmed that AL tends to select many samples from the hardest classes to discriminate among the others, as demonstrated in Joshi et al. (2009) and Di and Crawford (2011). The good performance of random sampling can easily be explained by the high probability of the random process finding new classes only by chance (Tuia et al., 2011a). From a learning point of view, learning tasks differ in this case because the label spaces between domains are not exactly the same (Pan and Yang, 2010).

5.5 Mixed classes confuse the selection process with active learning

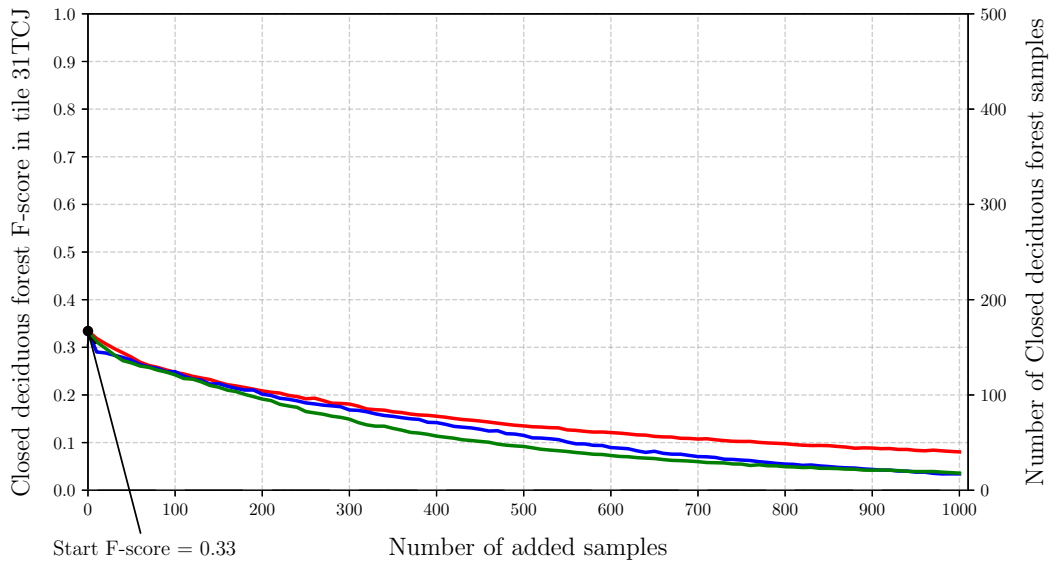
In all reported results (except those in Section 4.2.5), pixels of closed and open deciduous forests (mixed classes) were only used for testing but not for training. In this way, we made AL more efficient, as explained below.

In the digital forest maps used as sample sources, forest stands are defined as mixed species when the dominant species cover less than 75% of the total extent. Inside these stands, species can be spatially aggregated or scattered. Therefore, stands of mixed species can be composed of mixels (containing spectral information of several species in one pixel) or can be a mixture of pure pixels. If mixed-species classes are included in the training set, two issues may arise: (i) pure-species pixels may be confused with pixels labelled as mixed species since these pixels can also be pure at this level (the label making sense at the stand level); (ii) predictions of mixed-species classes on another S2 tile may be inefficient because of sample selection bias as the composition of mixed-species classes can vary across space from one tile to another although this can be corrected by AL. We observed the first issue for several species with confusions between pure and mixed-species classes.

Taking the example of the locust class, at the beginning of the transfer from the north-eastern source tile (31UEQ) to the south-western target tile (31TCJ), the locust F-score was very low with no additional samples. This poor performance was surprising because pure pixels of locust species were present in both source and target tiles. Due to spectral and temporal overlaps (see Figures A.5, A.6 and A.7), there was considerable confusion with oak and closed deciduous forest classes (Figure A.2b).



(a) Training with mixed classes



(b) Training without mixed classes



Figure 5.12 – Changes in the average F-score values on the south-western target tile (31TCJ) for the closed deciduous forest class according to the number of additional samples. The initial classifier was trained on the north-eastern source tile (31UEQ). In (a), the training set contained pure and mixed species while in (b) only pure classes were considered (explaining why no extra pixels of closed deciduous forest were added).

Considering mixed-species classes for training, the addition of extra target samples by AL led to a rapid increase in the locust F-score, especially by AL_{MS} but the performance boost was hampered by the closed deciduous forest class. Despite the low initial locust F-score, AL did not only select locust samples but also queried many uncertain samples belonging to the closed deciduous forest class (see Figure 5.12a). After checking, some of the uncertain pixels

turned out to be locust pixels (the other ones being pixels of oak).

Thus, AL seeks to improve the classification of locust in both cases but pixels can be found in two different classes. The existence of mixed-species classes makes AL less competitive (see Figures 5.9a and 5.9b for comparison). Besides, by selecting locust pixels labelled as closed deciduous forest in mixed stands, the model was progressively distorted. The negative effect of mixed-species classes was particularly pronounced with entropy-based AL. Because this metric takes the probabilities of belonging to all the classes into account (not only the two most probable classes, as in AL_{MS} which are locust and closed deciduous forest or oak), the model was thus influenced by low probabilities of unimportant classes which are rarely the right ones (Joshi et al., 2009). In other words, AL_H is less robust to class overlap. Regarding the model based on random sampling, more locust samples were selected than by AL_H which further improved the performance for this class. The same behaviour was encountered in Tuia et al. (2011a), where random sampling outperformed active learning in a similar context with highly mixed crop classes.

With pixels of pure-species classes only, the model performance improved for the locust class, especially with few extra samples. The decrease in the closed deciduous forest F-score (Figure 5.12b) can be explained by a better definition of the pure classes in the model. Pure pixels belonging to the closed deciduous forest class are rather identified in their corresponding pure classes. The model not only improved the definition of pure classes but also allowed a better characterisation of the species in the mixed-species stands, opening interesting perspectives to enrich the forest reference map.

6. Conclusion

In this paper, we propose the use of an active learning approach for the classification of poplar plantations, among other tree species, in an operational context. The results show the good capability of Sentinel-2 time series to map poplar plantations with all their diversity related to cultivars, management practices and climatic conditions. In particular, we demonstrate the potential of active learning to adapt a local model into a global model suitable for a national scale mapping. By adaptability, we have not only shown its efficiency to query relevant samples from the unexplored areas, but also its capacity to discover new classes.

Following the analysis of the different transfer combinations by active learning between the three Sentinel-2 study tiles, we conclude that spatial transferability is strongly dependent on the transfer direction but the performance gap tends to decrease with three-tile-based models. The resulting global models achieved high classification performances with few training samples reflecting the potential of active learning to considerably reduce the cost of labelling. With respect to untested uncertainty metrics, margin sampling provided the best classification

results and proved to be more robust to class overlap than entropy. We therefore recommend the use of the margin sampling metric for multi-class tasks, especially when dealing with mixed classes.

The proposed approach is not based on the algorithm properties as may be the case with some active learning settings using SVM that relies on algorithm-specific heuristics to obtain the support vectors. It could potentially be employed with other classifiers which provide class posterior probabilities either directly inferred using Bayes' rule or estimated for non-probabilistic classifiers. However, it is noteworthy to mention that, for the same labelling budget, the performance of active learning may vary depending on the classifier involved. Ensemble-based active learning approaches could potentially be considered as they ensure a robust selection while being independent of the classifier.

Other works with real-world applications would benefit from the use of active learning to guide field surveys or photo-interpretation operators in collecting labels only for relevant references rather than providing a pool of samples from which to choose the most useful ones.

5.3 Bilan

Dans ce chapitre, nous avons évalué l'apport de l'adaptation de domaine en adaptant un classifieur par apprentissage actif pour un transfert spatial des connaissances. Nous avons montré le potentiel de l'apprentissage actif pour adapter un modèle local spécifique à un site donné, à un modèle global capable de reconnaître des plantations de peupliers dans les différents contextes nationaux.

L'analyse des différentes combinaisons de transfert entre les trois tuiles Sentinel-2 a montré que la capacité de transfert spatial est liée à la direction dans laquelle le transfert est effectué, et qu'avec un modèle global incluant la variabilité des trois tuiles avec un minimum d'échantillons, les meilleures performances ont été atteintes reflétant ainsi le potentiel de l'apprentissage actif à minimiser le coût en échantillons d'entraînement étiquetés.

Cette approche reste toutefois dépendante du critère de choix des échantillons de référence à étiqueter, qui doit être choisi avec précaution, notamment en cas de chevauchement entre les classes.

Malgré la pertinence de l'apprentissage actif pour la cartographie à grande échelle et son intérêt à réduire le coût en termes d'échantillons de référence tout en garantissant de bonnes performances de classification, la mise en place de cette approche à l'échelle nationale semble compliquée. En effet, dans ces travaux nous avons restreint le choix de l'apprentissage actif des échantillons à étiqueter à eux qui sont disponibles dans la base d'échantillons de référence. En réalité, l'algorithme est censé explorer toutes les régions à classer et identifier ensuite les échantillons les plus informatifs. Si une telle approche doit être appliquée à l'échelle nationale, un temps de calcul important sera nécessaire avec le risque que les échantillons identifiés soient difficiles à étiqueter. De plus, la grande dimension spectrale et temporelle des séries temporelles Sentinel-2 rendra l'évaluation des échantillons encore plus fastidieuse.

Quatrième partie

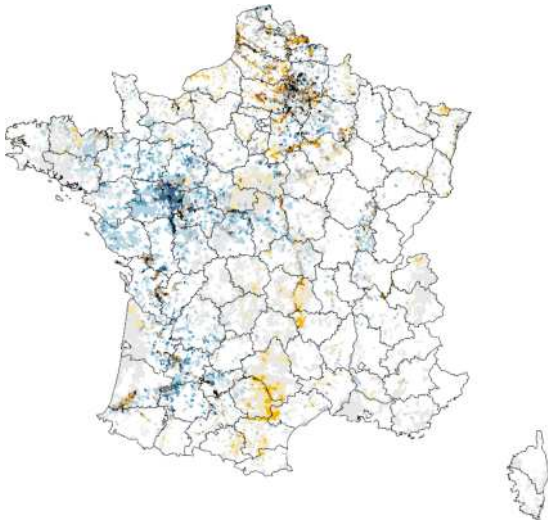
Vers une classification opérationnelle des peupleraies à l'échelle nationale

Chapitre 6

L'indice du peuplier pour la cartographie opérationnelle des peupleraies

Sommaire

| | | |
|-----|--|-----|
| 6.1 | Résumé | 147 |
| 6.2 | Sentinel-2 Poplar Index for Operational Mapping of Poplar Plantations over Large Areas | 148 |
| 1. | Introduction | 149 |
| 2. | Materials | 151 |
| 3. | Methods | 154 |
| 4. | Results | 160 |
| 5. | Discussion | 169 |
| 6. | Conclusions | 174 |
| 6.3 | Bilan | 176 |



6.1 Résumé

Ce chapitre présente la troisième approche proposée dans le cadre de ces travaux de thèse afin de répondre au problème d'identification des peupleraies sur de grandes étendues à partir des données Sentinel-2.

Dans cette partie, nous cherchons à identifier les caractéristiques spectrales et temporelles permettant de distinguer les peupleraies des autres essences feuillues. Ainsi, nous comparons plusieurs configurations établies à partir de bandes brutes et d'indices spectraux sur deux années de séries temporelles, 2017 et 2018, afin de déterminer les régions spectrales optimales et leurs périodes correspondantes qui sont les plus pertinentes pour les peupleraies. À partir des bandes spectrales les plus discriminantes, nous définissons également un nouvel indice de détection des peupliers - le *Poplar Index* - avec quatre variantes.

Les résultats révèlent la pertinence du moyen infrarouge (*Short Wave Infrared* (SWIR)) suivi de la région du *Red edge* au cours de la période printemps-été pour détecter avec précision les peupleraies, traduisant la sensibilité des peupliers à la teneur en eau tout au long de leur cycle phénologique.

Les meilleures performances avec les résultats les plus stables ont été obtenues avec l'indice du peuplier PI_2 combinant les bandes spectrales B5, B11 et B12. L'indice PI_2 a été validé sur les deux années et utilisé pour produire la carte nationale des peupleraies en 2018 avec une précision producteur moyenne de 95%.

Cette étude fournit une approche opérationnelle pour le suivi de la ressource en peuplier sur de grandes étendues par les gestionnaires forestiers.

Mots-clés : Populus; Sentinel-2; Indice Spectral; Sélection de variables; Essences d'arbres; Classification d'images

6.2 Sentinel-2 Poplar Index for Operational Mapping of Poplar Plantations over Large Areas

Authors: Yousra Hamrouni, Eric Paillassa, Véronique Chéret, Claude Monteil, David Sheeren

Paper status: Submitted to Remote Sensing of Environment on 19 June 2021.

Abstract: Poplar (*Populus* spp.) is a fast-growing tree planted to meet the growing global demand for wood products. In France, the country with the largest area planted with poplar in Europe, accurate and up-to-date maps of its spatial distribution are not available at the national scale. This makes it difficult to estimate the extent and location of the poplar resource and calls for the development of a robust and timely stable approach for mapping large areas in order to ensure efficient monitoring. In this study, we investigate the potential of Sentinel-2 time series to map the diversity of poplar plantations at the French countrywide scale. By comparing multiple configurations of spectral features based on spectral bands and indices over two years (2017 and 2018), we identify the optimal spectral regions with their respective time periods to distinguish poplar plantations from other deciduous species. We also define a novel poplar detection index (PI) with four variants that combine the best discriminative spectral bands. The results highlight the relevance of SWIR followed by red edge regions in the spring-summer period to accurately detect poplar plantations, reflecting the sensitivity of poplar trees to water content throughout its phenological cycle. The best performances with stable results were obtained with the PI₂ poplar index combining B5, B11 and B12 spectral bands. The PI₂ index was validated over two years and used to produce the national map of poplar plantations in 2018 with an average producer's accuracy of 95%. This study provides an operational approach for monitoring the poplar resource over large areas by forest managers.

Keywords: Populus; Sentinel-2; Spectral Index; Feature selection; Tree species; Image classification

1. Introduction

Timely and accurate mapping of the distribution and composition of forest stands is essential for both forest management and biodiversity assessment (Fares et al., 2015). According to the most recent Global Forest Resources Assessment by the Food and Agriculture Organization (FAO), the total area of the world's forests is 4.06 billion hectares of which, 93% are naturally regenerating forests and 7% are planted forests (FAO, 2020). While the area of naturally generating forests has been decreasing since 1990, planted forests have increased by about 123 million hectares over the same period, indirectly compensating for some of the area lost through deforestation (Carle and Holmgren, 2008). The importance of planted forests has continued to increase as wood has become an increasingly crucial raw material for industry and an attractive source of bioenergy. In the last two decades, short-rotation coppice (SRC) plantations such as poplar are among the most important fast-growing planted tree species. Poplar plantations are widely distributed throughout the world and mainly cultivated in China, western Europe and North America for their high quality timber, high growth rate and short rotation cycle of about 20 years (FAO, 2016a).

In France, poplar plantations are an important economic resource mainly exploited to satisfy the growing demand for wood. According to the latest report by the International Poplar Commission, the surface area of poplar plantations in France was estimated at 0.2 million hectares (FAO, 2016a), i.e., one tenth of the national forest. These area estimates come from forest databases, cadastral surveys, or statistical forest inventories, but have never been established synchronously using a homogeneous approach throughout the country, and the extent and location of the plantations are generally not reported. Consequently, it is difficult to accurately estimate the surface area of these plantations and regularly monitor the dynamics of the poplar resource.

Different remote sensing techniques have been developed to map forest stands (Boyd, 2005) by identifying forest cover and the dominant leaf type e.g., broadleaf or coniferous (Reitberger et al., 2008; Hansen et al., 2020) and even go as far as characterising the composition of forest tree species (Immitzer et al., 2012; Sheeren et al., 2016; Fassnacht et al., 2016; Persson et al., 2018).

Some studies focussed more specifically on mapping forest plantations (Torbick et al., 2016) notably rubber (Li and Fox, 2012; Ye et al., 2018; Xiao et al., 2019), palm oil (Puttinaovarat and Horkaew, 2019; Descals et al., 2019) and eucalyptus (le Maire et al., 2014; Forstmaier et al., 2020; Deng et al., 2020), but also acacia (Spracklen and Spracklen, 2021). Although significant efforts have been made to identify some natural poplars (mainly European aspen, *Populus tremula* L. and quaking aspen, *Populus tremuloides* Michx.) (Heyman et al., 2003; Viinikka et al., 2020; Kivinen et al., 2020), little effort has been devoted to poplar plantations

in temperate regions, including the many different cultivars used and stand development stages (Chardenon and Flouzat, 1981; Borry et al., 1993; Grignetti et al., 2009; Eslami and Zahedi, 2011).

Chardenon and Flouzat (1981) tested the capabilities of Landsat-2 imagery to classify poplar stands in France (in the Garonne River plain) and Italy (in the Pô valley), with different canopy cover rates related to stand age. These authors achieved only limited success using a single image acquired in May, July and October 1972 and in June, July, and September 1975. Discrimination was improved using a bitemporal classification based on images acquired in June and September 1975 with 49.4% accuracy for young plantations (cover rate < 25%) and 80.8% for older (cover rate > 25%). Young plantations were difficult to identify due to the presence of herbaceous understory vegetation. The authors also remarked on the inadequacy of the Landsat-2 spatial resolution (80 m) to identify many small poplar stands (< 1 ha). In another study, Borry et al. (1993) concluded that the discrimination of poplar stands at different development stages is possible using one single image acquired in the full growing season. Their results, in contrast to those of Chardenon and Flouzat (1981), were obtained with SPOT-1 HRV (20 m spatial resolution) and Landsat-5 TM (30 m) images. Based on a per-field classification, young stands were classified with 80-94% accuracy, middle age stands with 45-77% accuracy, and older with 46-74% accuracy depending on the spectral band and the date selected. The authors found combining several dates did not improve the discrimination. The stand sizes were relatively small, ranging from 0.5 to 2 hectares, with different types of understorey vegetation (herbaceous layer, dense or scattered shrubs). Object-based classification of poplar plantations using very high spatial resolution (Ikonos and Quickbird) imagery was also investigated by Grignetti et al. (2009). Overall accuracy ranging from 74% to 82% was obtained but with very unreliable identification in the case of new plantations (< 3 years old). Other interesting approaches at tree level and stand level are reported in Kivinen et al. (2020) based on LiDAR data, airborne photogrammetry, multispectral and hyperspectral imaging. However, most concern aspen trees in natural contexts, especially in boreal forest landscapes.

The above-mentioned studies produced contrasting results concerning the ability to produce accurate maps of poplar plantations in various management contexts. Differences in the levels of reflectance between plantations of poplar species and other tree species in natural forests were not examined, nor was the possible identification of specific wavelengths regions or spectral indices for poplar discrimination. In addition, it is not clear whether integrating the temporal dimension in the classification procedure is useful even though seasonal variations in vegetation (i.e. phenology) could improve performances (Hościło and Lewandowska, 2019; Karasiak et al., 2019). Moreover, previous studies were limited to local areas making it

difficult to evaluate if identical findings would be obtained for larger areas.

In this study, we investigated the automatic classification of all poplar plantations at the French national scale using optical Sentinel-2 (S2) time series. We hypothesised that time-series data would enable better classification than mono-temporal images. Our main objective was to define a generic methodological approach to map the wide range of poplar plantations in an operational context. After a preliminary spectral analysis based on visual interpretation, we used a feature selection technique to select optimal bands and spectral indices at specific dates for poplar discrimination. We also defined a new S2 poplar detection index (PI) with four variants to improve the classification. We evaluated its stability over two years on three contrasting poplar sites at different stand development stages. The best version of PI was then used to produce the map of poplar plantations at the national scale. This study provides a practical approach for monitoring the poplar resource in the near future by plantation managers.

2. Materials

2.1 Satellite data

In collaboration with forest partners, we chose three Sentinel-2 image tiles (each covering 100 km x 100 km) to conduct the analysis. These tiles are located in the north-east, centre and south-west of France (Figure 6.1). They cover the most contrasting poplar sites in terms of diversity related to silvicultural practices, types of cultivars planted and climate conditions. The 31UEQ tile is located in the degraded oceanic climate region of the north and centre plains. The 30TYT tile is located in the altered oceanic climate while the 31TCJ is characterised by the southwest basin climate with possible summer drought (Joly et al., 2010).

Two years, 2017 and 2018, were considered independently to check the concordance of the results between years. All available S2 images of the three tiles and for the two years were downloaded from the French THEIA Land Data Center. Top Of Atmosphere (TOA) surface reflectance products (level 2A) were used, meaning the data have been orthorectified and corrected for topographic and atmospheric effects. Images were pre-processed by THEIA with the MAJA platform and a cloud mask was provided with each acquisition (Hagolle et al., 2015).

We only selected the 10-m and 20-m S2 spectral bands, the other bands (1, 9 and 10) at 60-m spatial resolution being dedicated to atmospheric correction and cloud screening. At 10 meters, four spectral bands are available: Blue (B2 - 490 nm), Green (B3 - 560 nm), Red (B4 - 665 nm) and Near Infrared (B8 - 842 nm). At 20 meters, six spectral bands are provided: Red edge (B5 - 705 nm, B6 - 740 nm, B7 - 783 nm), Near Infrared (8A - 865 nm), and Short-Wave



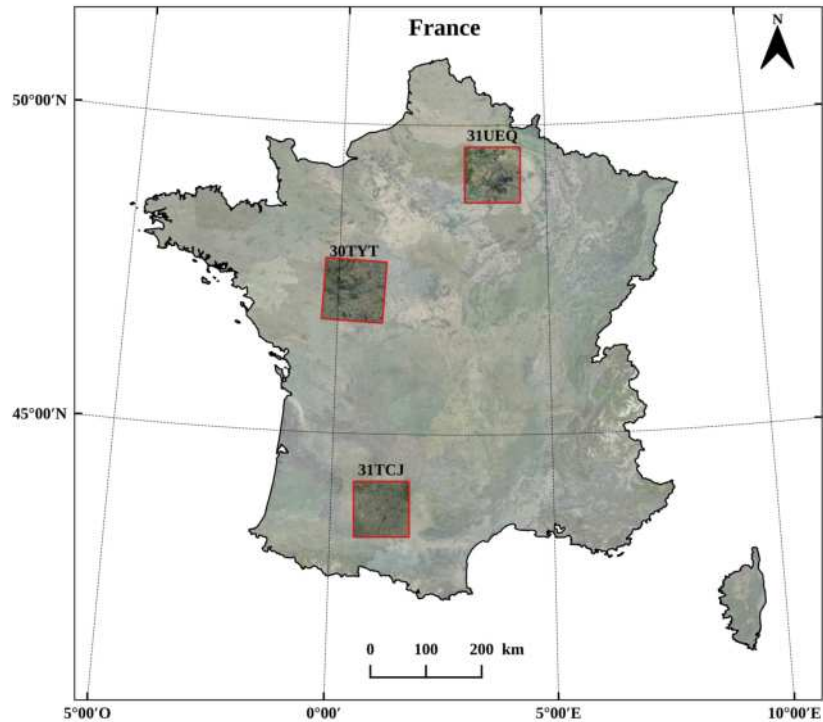


Figure 6.1 – Sentinel-2 image tiles used in this study: 31UEQ, 30TYT and 31TCJ. The extent of each tile is displayed as a red square with 100 km sides. S2 tiles are shown in the UTM/WGS84 projection.

Infrared (B11 - 1610 nm, B12 - 2190 nm).

Because of specific sensor constraints (e.g. different relative orbits) or the cloud cover, the acquisitions of one year are non-synchronous in the three study tiles. Therefore, for each S2 time series of the same year, all the images were resampled to 10 m spatial resolution with nearest-neighbour resampling and with the same 10-day time step for all tiles. While performing the temporal resampling, missing data (i.e. pixels that were invalid due to clouds or cloud shadows) were filled using a multi-temporal linear interpolation. This involved replacing each cloudy and shady pixel (detected in the cloud mask) with an interpolated value derived from the nearest valid pixels in the times series (Inglada et al., 2016). The resulting gap-filled image time series comprise 340 features in 2017 (34 dates x 10 spectral bands) and 360 features in 2018 (36 dates x 10 spectral bands). The resampled dates are detailed in Appendix B.1.

2.2 Reference data

Reference polygons for training and testing were drawn from the latest release of the French National Forest Inventory spatial database (BD Forêt@IGN, v.2) produced by the national mapping agency (IGN). This database provides a vector map of forest stands (polygons) with a minimum area of 0.5 hectares. The composition of each stand was obtained by interpreting aerial stereo-images completed by field surveys. This map was produced progressively, district by district, starting in 2007. National coverage was completed in 2018. We only considered

classes of deciduous species in this forest map (including poplar plantations). Coniferous species were excluded. Pure and mixed stands of deciduous species were retained. While pure stands consist of polygons covered by 75% of a single dominant species, mixed stands have only 50% to 75% coverage. In the database, poplar plantation polygons are always referred to as pure stands.

Since there is a time lag between the year when the forest database was made available (between 2007 and 2018 depending on the district) and the year of acquisition of the S2 time series, all poplar polygons for both 2017 and 2018 were checked through photo-interpretation by three experts. This inspection step ensures up-to-date references as poplar stands may change over time due to their short rotation cycle. We selected references of various stand ages and cultivars but very young plantations (< 3 years approximately) were excluded because of their limited canopy cover rate and the difficulty involved in identifying them accurately through photo-interpretation.

A set of poplar polygons from the southwestern tile (31TCJ) was validated in the field by forest partners. Out of a total of 230 photo-interpreted polygons, two were eucalyptus plantations, representing less than 1% misidentification. Another set of 85 polygons from a field campaign was also used to identify the limit of detection of poplar stands according to their stage of development. Reference pixels for the other deciduous species were drawn directly from the forest database, without checking or updating.

A data-cleaning step was applied to all the pixels included in the reference polygons of the three S2 tiles to eliminate outliers arising from undetected clouds in the time series (i.e. for which no gap-filling had been performed) or from pixels that differ significantly from their membership class (e.g. urban pixels or pixels showing bare soil in pure stands). The outlier detection procedure followed the 1.5 interquartile range rule (1.5 IQR), which considers values outside the interval $[Q1 - 1.5 IQR ; Q3 + 1.5 IQR]$ as outliers, where $Q1$ and $Q3$ denote the first and third quartiles respectively. The outlier detection procedure first consisted in eliminating all pixels outside the predefined interval in each band of the time series and for each tree species class separately. Thereafter, and proceeding by class, only the valid pixels common to all bands in the time series were retained. In this way, if pixels were detected as outliers in only one band, they were automatically removed from the sample set. Finally, the number of samples per class was set so that the smallest class was fully sampled.

The total number of sampled pixels in each tile and for each year, after removal of the outliers, is listed in Table 6.1. The sample size for poplars varied between 2,500 and 7,700 pixels in 2017 and between 3,200 and 5,000 in 2018. The difference between the number of samples in 2017 and 2018 is due to the fact there were more outliers in 2018. The locust tree class was removed from tile 31UEQ in 2018 because the majority of its samples were identified

as outliers and the number of remaining pixels was too small to be considered representative of the whole tile. This was also the case for other species.

Table 6.1 – Total number of reference samples per class extracted from each S2 tile, after removal of outliers.

| Sample size in pixels per class | | | | | | | | |
|---------------------------------|-----------|--------|-----------------|-----------------|------|-----------------|---------------------------------|-------------------------------|
| Year | Tile code | Poplar | Locust tree | Chestnut | Oak | Beech | Closed deciduous forest (mixed) | Open deciduous forest (mixed) |
| 2017 | 31UEQ | 2500 | 2500 | NA ¹ | 2500 | 2500 | 2500 | 2500 |
| | 30TYT | 4000 | 4000 | 4000 | 4000 | NA ¹ | 4000 | 4000 |
| | 31TCJ | 7700 | 7700 | 7700 | 7700 | NA ¹ | 7700 | 7700 |
| 2018 | 31UEQ | 3700 | NA ¹ | NA ¹ | 3700 | 3700 | 3700 | 3700 |
| | 30TYT | 3200 | 3200 | 3200 | 3200 | NA ¹ | 3200 | 3200 |
| | 31TCJ | 5000 | 5000 | 5000 | 5000 | NA ¹ | 5000 | 5000 |

¹ Class not available in the study area or poorly represented.

3. Methods

We used a statistical approach based on the sequential forward floating selection (SFFS) algorithm to measure the importance of spectral features and to select a subset of bands and spectral indices for the classification of poplar stands among the other tree species. The analysis was conducted in three steps. First, we detected the most relevant spectral regions and time periods using a combination of all S2 spectral bands and acquisitions of the time series (Figure 6.2a). This was carried out separately for 2017 and 2018. Then, based on these results, we identified a set of existing potentially discriminative spectral indices. We also defined a new S2 poplar detection index (PI) with four variants to improve the classification. We selected features from this set of data to identify the most relevant combination of spectral indices (Figure 6.2b). Finally, we evaluated each spectral feature independently using all available dates in each time series (Figure 6.2c and 6.2d) and compared the classification performances with those of the previous subsets of features (combination of spectral bands or indices). The best feature subset with the most stable performance for the two years was used to produce the map of poplar plantations at the national scale.

3.1 Feature selection with SFFS

Several techniques have been proposed to address the problem of identifying a subset of relevant variables (features) in a large input dataset (Jia et al., 2013; Chandrashekar and Sahin, 2014). In remote sensing, this helps to identify specific spectral features and periods related to the classes of interest, before examining the possible origins and before trying to understand the underlying physical processes. This also reduces the effect of the curse of

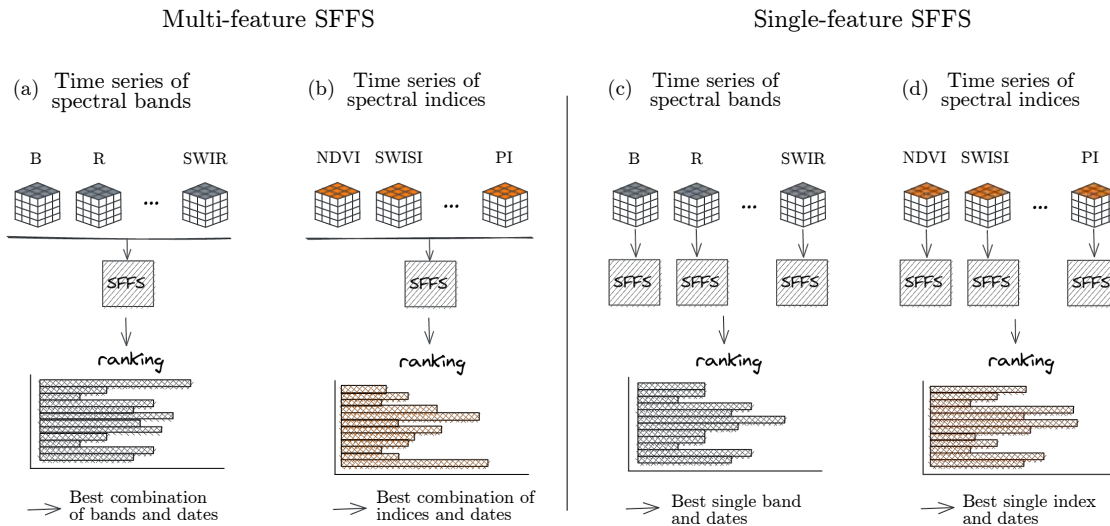


Figure 6.2 – Single and multi-feature SFFS configurations to measure the subsets of spectral features that are the most relevant for mapping poplar stands.

dimensionality which has a negative impact on classification performance. In addition, it eliminates redundant or noisy features and reduces the computation time required to process the data.

In this study, we used SFFS, a wrapper-based feature selection algorithm (Pudil et al., 1994). SFFS is a standard method that has already been shown to provide a good trade-off between computation time and efficiency (Gomez-Chova et al., 2003; Le Bris et al., 2014). Compared to feature extraction techniques that rely on data transformation methods (such as principal component analysis) independently of the classification, wrapper-based approaches perform feature selection using a classification algorithm to evaluate subsets, the predictive performance being the objective function (Chandrashekar and Sahin, 2014).

SFFS is a bottom-up search procedure which selects an optimal subset of features, starting from an empty set, and at each iteration, adds the most important feature (i.e. the one with the highest value for the objective function) to the pool of selected features. SFFS is a more flexible version of the initial sequential forward selection (SFS) method in which individual features are permanently conserved after selection, with no backtracking. The floating search approach overcomes the so-called *nesting effect* by performing a backward selection after each SFS step to conditionally exclude one of the previously selected features, thus avoiding a local minimum (Pudil et al., 1994). The feature is removed if a subset of the same size achieves better accuracy according to a given performance criterion. In this way, it is possible to rectify the inclusion of previous features and to replace them with new better features in order to converge towards the optimal solution. The iteration continues until the required number of features is added or the expected performance is reached.

This method was used to identify the most relevant subsets that offer the best classific-

ation performance in each feature space (i.e. for both multi-feature and single-feature SFFS configurations, Figure 6.2). This was carried out until ten features were included in the subset.

3.2 Sample selection, classification and comparison of performances

Reference polygons (forest stands) were randomly separated into 50% for training and 50% for testing. The S2 pixels (10 m square) were then sampled in each subset of polygons and feature values were extracted from each gap-filled time series in 2017 and 2018. To ensure the balance of classes in each tile, training and testing pixels were selected from reference polygons using stratified random sampling of equal size for each class. Training and testing pixels were collected from spatially disjoint reference polygons in order to limit spatial autocorrelation effects (Karasiak et al., 2021).

SFFS was computed only on the training set (50% of the samples) with five-fold cross validation. Each model with the best subset of ten features (one for each SFFS configuration) was then evaluated on the independent test set (50% of the remaining samples).

We used random forest (RF) as a supervised classification algorithm (Breiman, 2001). Tuning the hyperparameters, including the number of decision trees, the maximum depth of the trees and the number of features considered by each tree when splitting a node, was based on a grid-search strategy and five-fold cross validation. The number of trees tested ranged from 10 to 150 with a step of 10. The values tested for trees depth included the maximum possible value that expands the tree nodes until all the leaves become pure as well as intermediate values between 5 and 50 with an interval of 10. As recommended, the number of features to split each node was set to the square root of the total number of input variables (Breiman, 2001; Ghosh and Joshi, 2014) in addition to a range of values between 1 to 20 with a step of 2.

Classification performance was assessed in terms of producer's accuracy (PA), also known as recall, for the poplar class. PA was used as the objective function in SFFS. It is a non-prevalence-dependent performance measure that, for a given class, corresponds to the number of correctly identified samples among all reference samples assigned to that class. As recommended by Foody (2020), we deliberately disregarded the kappa metric.

The performances of the best models (i.e. those built with the optimal feature subsets derived for each configuration in Figure 6.2) were compared using the McNemar's statistical test (McNemar, 1947) which is a non-parametric alternative to the T-test to examine the statistical significance of performance differences between two classifiers (Dietterich, 1998). McNemar's test is based on a 2x2 contingency matrix relying on a binary distinction between correct and incorrect class assignments (Foody, 2004). The McNemar test was applied to compare each pair of models separately under the null hypothesis that their predictive performances were

equal with a significance level of $\alpha = 0.05$.

The feature selection, classification and performance evaluation were implemented in Python using the *mlxtend* (Raschka, 2018) and *scikit-learn* (Pedregosa et al., 2011) libraries.

3.3 Sentinel-2 Poplar detection index (PI)

Prior to statistical processing, the spectral profiles were analysed visually and changes in poplar profiles over time evaluated by comparing the poplar reference samples and the other deciduous species (Figure 6.3). This analysis revealed a specific spectral pattern for poplars: from mid April to early September, an absorption valley appeared in the SWIR bands B11 and B12 with no overlap with the other tree species (Figure 6.4). The reflectance of the SWIR bands decreased for all species during this period but the decline was more pronounced for poplar plantations, especially in B11. This pattern was observed in both the 2017 and 2018 S2 time series.

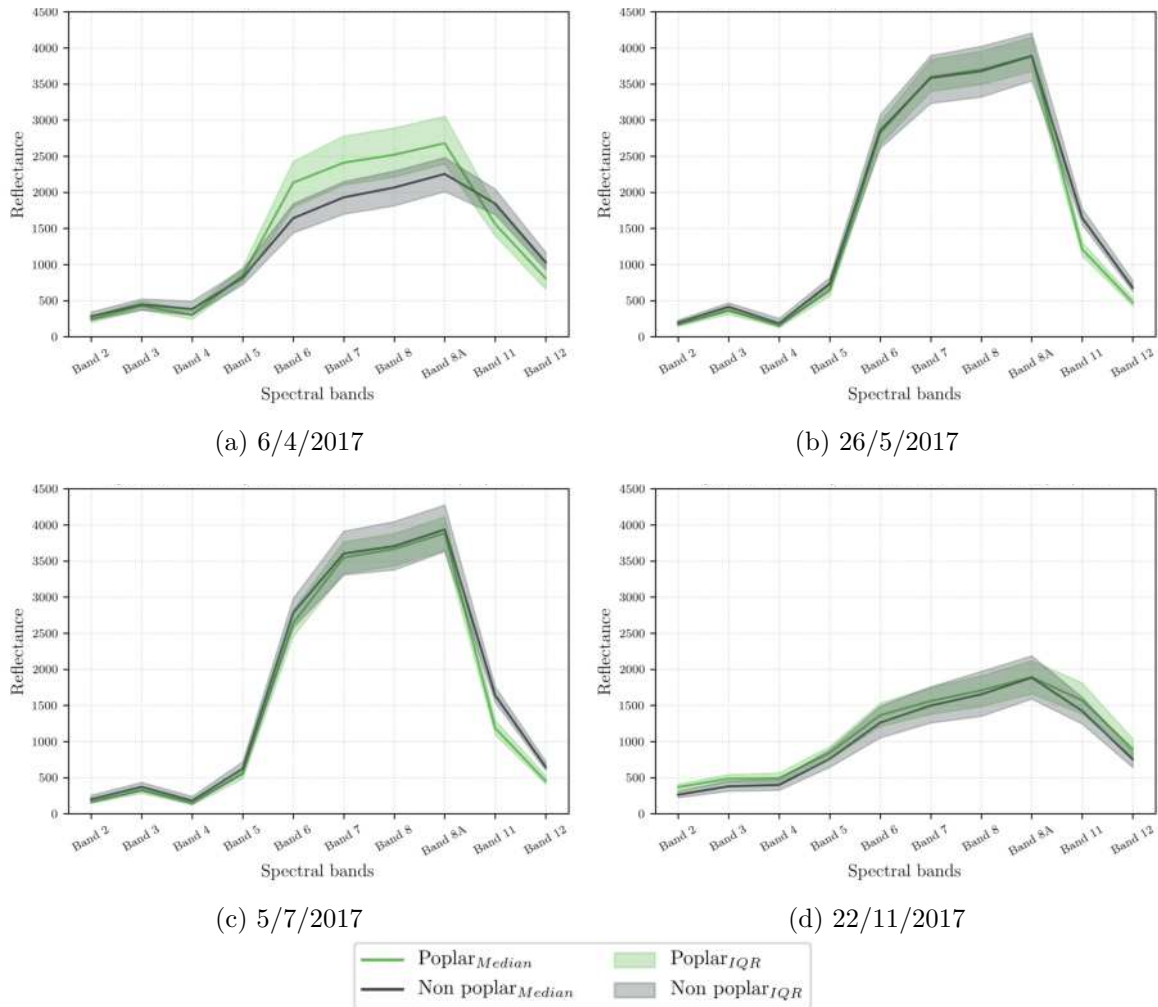


Figure 6.3 – Spectral reflectance (multiplied by a factor 10,000) of poplar plantations versus deciduous tree species on four different dates in 2017. The solid lines represent the median signal and the surrounding shaded areas represent the interquartile range, in green for poplar and in grey for the other deciduous species.

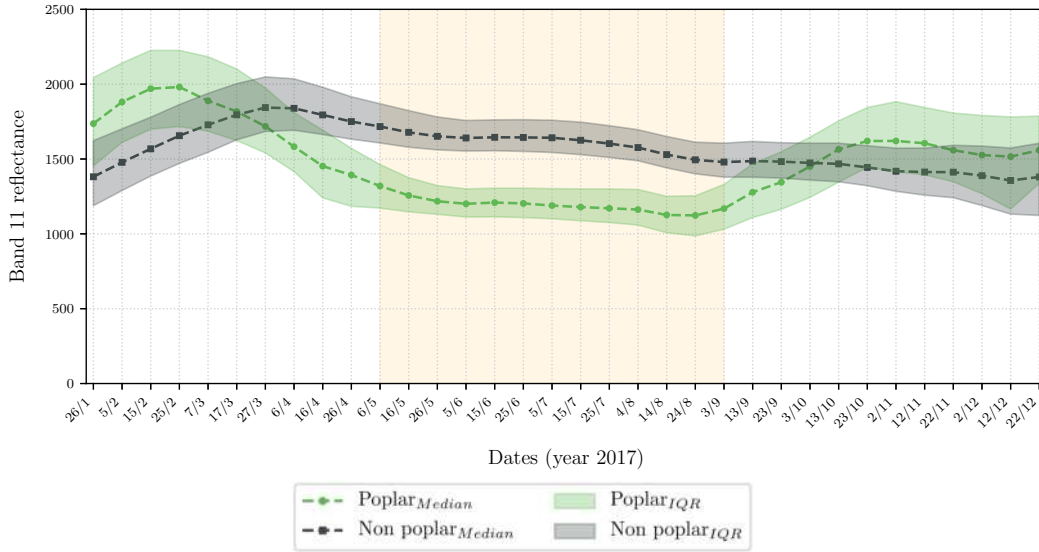


Figure 6.4 – Temporal profile of Band 11 reflectance (multiplied by a factor 10,000) in poplar plantations versus deciduous tree species in 2017. The dashed lines represent the median signal and the surrounding shaded areas represent the interquartile range, in green for poplar and in grey for the other deciduous species. The pale yellow box highlights the most discriminating spring-summer period for poplars.

Based on these observations, we established a new S2 poplar detection index (PI) with four variants (1, 2, 3, 4) as alternatives to spectral bands or existing spectral indices. Formulations are as follows:

$$PI_1 = B11 - B12 \quad (6.1)$$

$$PI_2 = B5 - (B11 + B12) \quad (6.2)$$

$$PI_3 = \frac{B5 - (B11 + B12)}{B5 + (B11 + B12)} \quad (6.3)$$

$$PI_4 = B11 + B12 \quad (6.4)$$

The PI_1 index is expressed as the difference between $SWIR_1$ (1610 nm) and $SWIR_2$ (2190 nm). This difference was more pronounced for poplar stands than for the other tree species. This was also true between $SWIR_1$ and other spectral bands (e.g. B8A, B8) but this difference was tested using other existing spectral indices (see below). The PI_2 index is defined as the addition of the two SWIR bands which is then, subtracted to the red edge band (B5 centered at 705 nm). The interest of this formulation is using two wavelength regions and a spectral band with very close reflectance values for poplar stands and other species, in addition to the SWIR bands B11 and B12. This formulation accentuates the gaps. A normalised version is

proposed with the PI_3 index. The most intuitive PI_4 is a simple addition of the SWIR bands, which are expected to be lower for poplar stands. However, because PI_4 and PI_2 were highly correlated (Spearman's $\rho = 0.95$ in 2017 and 0.93 in 2018; Appendix B.2) we excluded PI_4 from the multi-feature SFFS analysis. Results are only provided for the single-feature SFFS configuration.

These new indices were compared to existing spectral indices. Here, we only selected standard indices (e.g. NDVI) or indices including SWIR bands or red edge domains (Table 6.2). A more exhaustive list of indices was tested but for the sake of clarity, these less competitive indices are only reported in Appendix B.3.

A total of 16 spectral indices were computed for each S2 acquisition, resulting in 544 features in 2017 (i.e. 16 spectral indices x 34 dates) and 576 features in 2018 (36 dates). Using only spectral bands, the number of features was lower (340 and 360 for 2017 and 2018 respectively). In the single-feature SFFS configurations, the number of features for one band or one spectral index was directly linked to the number of acquisition dates.

Table 6.2 – List of the spectral indices compared with the corresponding formulas and references.

| Indices | Formulas | References |
|--|---|-------------------------------|
| Normalised Difference Vegetation Index | $NDVI = \frac{B8-B4}{B8+B4}$ | (Rouse et al., 1974) |
| Simple ratio Moisture Index | $MSI = \frac{B11}{B8}$ | (Hunt and Rock, 1989) |
| Simple ratio Disease Water Index 4 | $DSWI4 = \frac{B3}{B4}$ | (Apan et al., 2003) |
| Normalised Pigment Chlorophyll ratio Index | $NPCRI = \frac{B4-B2}{B4+B2}$ | (Peñuelas et al., 1994) |
| Normalised Burned Ratio Index | $NBRI = \frac{B8-B12}{B8+B12}$ | (Key and Benson, 2006) |
| Shortwave Infrared Water Stress Index | $SIWSI = \frac{B8A-B11}{B8A+B11}$ | (Fensholt and Sandholt, 2003) |
| Anthocyanin Reflectance Index | $ARI = \frac{1}{B3} - \frac{1}{B5}$ | (Gitelson et al., 2001) |
| Soil Adjusted Vegetation Index | $OSAVI = \frac{(1+0.16) \cdot (B8-B4)}{(B8+B4+0.16)}$ | (Rondeaux et al., 1996) |
| Leaf Chlorophyll Index | $LCI = \frac{B8-B5}{B8+B4}$ | (Datt, 1999) |
| Modified Chlorophyll Absorption in Reflectance Index | $MCARI = (B5-B4) - 0.2 \cdot (B5-B3) \cdot (\frac{B5}{B4})$ | (Apan et al., 2003) |
| Red edge Index 2 | Red edge2 = $\frac{B5-B4}{B5+B4}$ | (Cloutis et al., 1996) |
| SWIR ratio | SWIR ratio = $\frac{B12}{B11}$ | (Guerschman et al., 2009) |
| Poplar Index 1 | $PI_1 = B11 - B12$ | Equation 6.1 |
| Poplar Index 2 | $PI_2 = B5 - (B11 + B12)$ | Equation 6.2 |
| Poplar Index 3 | $PI_3 = \frac{B5 - (B11 + B12)}{B5 + (B11 + B12)}$ | Equation 6.3 |
| Poplar Index 4 | $PI_4 = B11 + B12$ | Equation 6.4 |

3.4 National mapping of poplar plantations

The best subset of features identified by SFFS was retained to map poplar plantations at the national scale from a full model. For this purpose, we used the *iota*² processing chain which is designed for operational land cover mapping at country scale and enables optimal

management of large volumes of satellite data (Inglada et al., 2017). *iota*² was configured to use S2 data of the 2018 time series. All the images were subjected to the same pre-processing steps including gap-filling, resampling of all bands to a 10 m spatial resolution and temporal resampling with a 10-day step. The RF classification model was trained with the whole set of available references in the three study tiles and was then used to predict the 90 S2 tiles that cover the whole of France. The French Space Agency (CNES) computing infrastructure was used to process and store all the data.

Predictions in the resulting national map were partially masked with the freely available pan-European Copernicus High Resolution Layers (HRL) 2018 forest layer of dominant leaf type at 10 m resolution (see <https://land.copernicus.eu/pan-european/high-resolution-layers/forests>) to only retain pixels of forest with their membership class. The surface area of each poplar plantation in each municipality was then computed and mapped with its median confidence value.

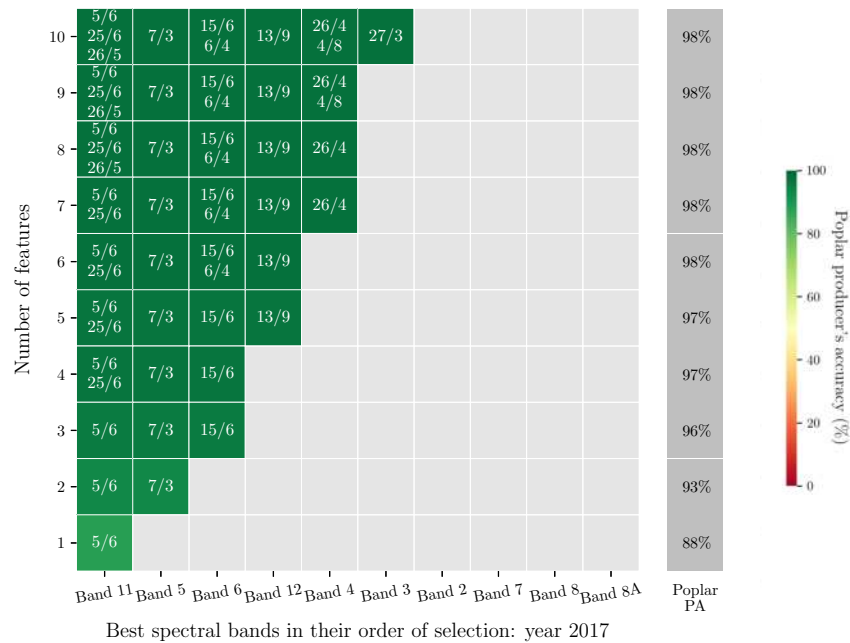
4. Results

4.1 Multi-feature SFFS

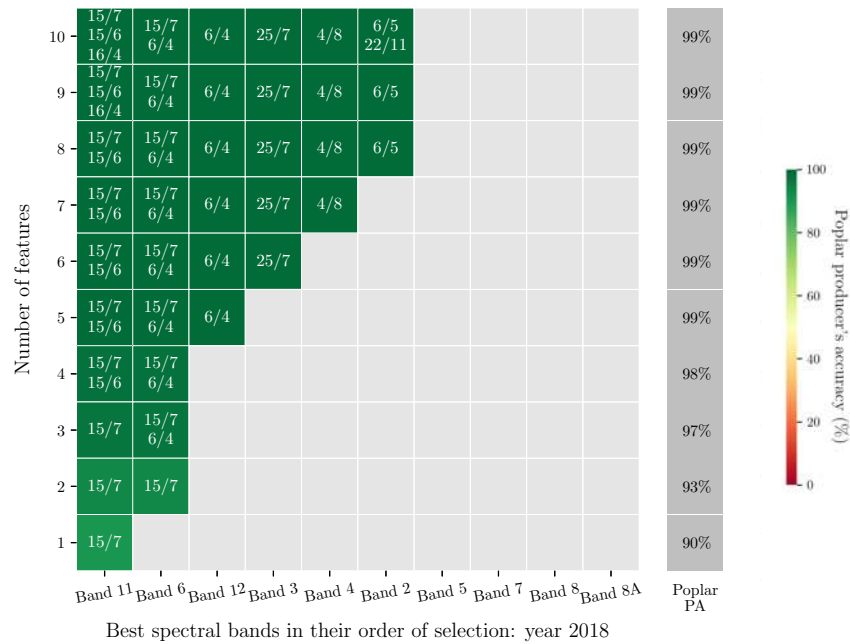
The results of feature selection with SFFS on the pool of spectral bands (i.e. multi-band configuration) are summarised in Figure 6.5. These graphs provide for each number of features selected, ranging from 1 to 10, the best performance achieved in terms of average poplar producer's accuracy (PA, based on 5-fold CV), and the best bands selected with their corresponding dates.

In both years, the best band selected with SFFS was B11 (SWIR spectral range), which reached poplar PA values of 88% in 2017 and 90% in 2018 using one date. The dates of this best single feature were the 5th of June in 2017 and the 15th of July in 2018. In both 2017 and 2018, the performance improved to respectively, 98% and 99% and then levelled off with the addition of six and five features belonging exclusively to the SWIR (B11 and B12) and red edge (B5 and B6) domains. Although the dates of the best feature subsets differed between the two years, the selection period was mainly from early spring to late summer. In addition, despite the fact that SFFS can exclude some selected features in the newly updated set due to the floating search approach, we observed that each time a feature was selected, it was no longer excluded from the subset.

When feature selection was based on the spectral indices (i.e. multi-index configuration), the best first features differed between the two years (Figure 6.6). PI_1 emerged for 2017 (Figure 6.6a) while PI_2 was first selected for 2018 (Figure 6.6b). Considering only these indices, the poplar PA reached 91% in 2017 and 93% in 2018, i.e., a difference of 3% to B11, which ranked first in the previous configuration. The dates of these first indices coincide with

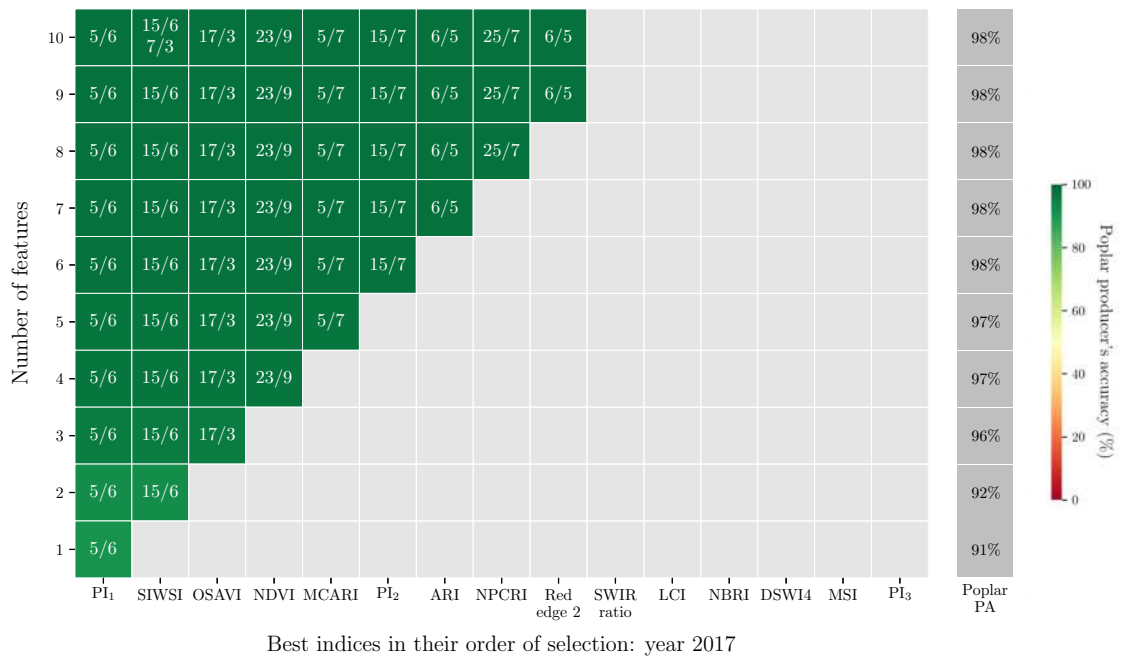


(a) Year 2017

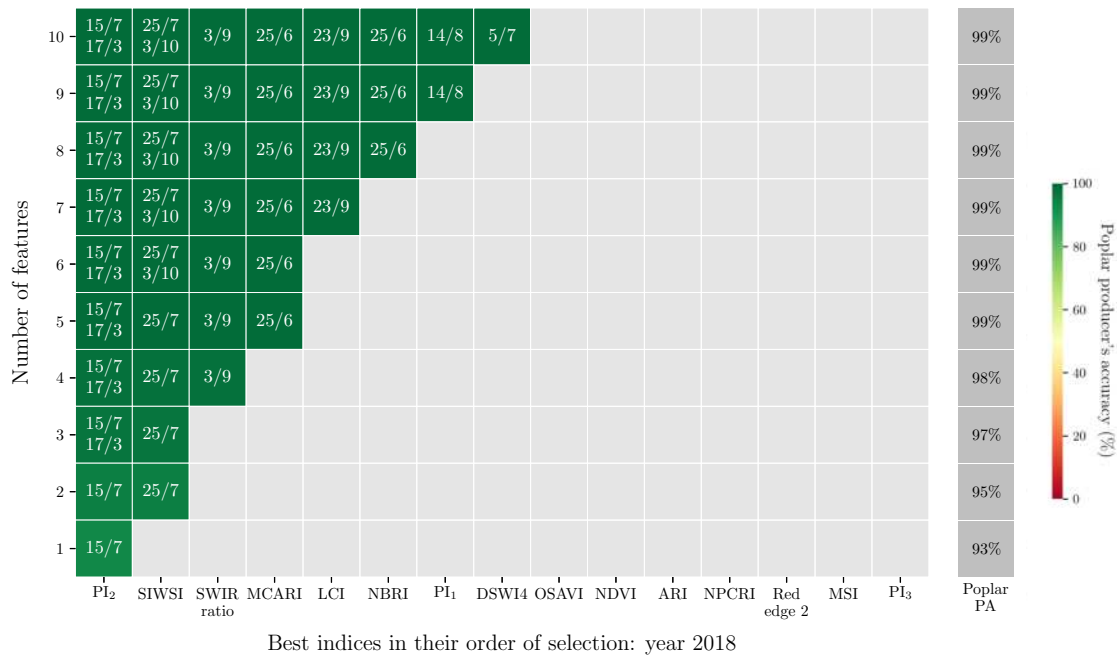


(b) Year 2018

Figure 6.5 – Average poplar PA as a function of the number of features selected with SFFS in the multi-band configuration. In (a) Multi-band SFFS results in 2017 and in (b) Multi-band SFFS results in 2018. The y-axis represents the number of features from 1 to a maximum of 10 and the corresponding spectral bands are listed in the x-axis in their order of appearance. The colours on the red-yellow-green scheme palette reflect the poplar's PA values, also displayed in front of each number of features (grey cells on the right). The dates of the selected bands are given inside the cells. If a cell contains more than one date, it implies that the band was selected several times on different dates. Empty cells in light grey indicate that the band in the x-axis was not selected for the given number of features on the y-axis.



(a) Year 2017



(b) Year 2018

Figure 6.6 – Average poplar PA as a function of the number of features selected with SFFS in the multi-index configuration. In (a) Multi-index SFFS results in 2017 and in (b) Multi-index SFFS results in 2018. The y-axis represents the number of features from 1 to a maximum of 10 and the corresponding spectral indices are listed in the x-axis in their order of appearance. The colours on the red-yellow-green scheme palette reflect the poplar’s PA values, also displayed in front of each number of features (grey cells on the right). The dates of the selected indices are given inside the cells. If a cell contains more than one date, it implies that the index was selected several times on different dates. Empty cells in light grey indicate that the index on the x-axis was not selected for the given number of features on the y-axis.

those of the best bands. The addition of other indices of different dates enabled maximum performance to be reached. Although the combination of indices differed in 2017 and 2018, some were the same (MCARI, SIWSI) in addition to PI_1 and PI_2 .

In these two multi-feature configurations involving all spectral bands or all spectral indices, the results highlight the interest of the SWIR and red edge domains, either by using the raw bands directly or by combining them in spectral indices. With a single feature, the multi-index configuration performed slightly better than the multi-band configuration. However, with the addition of new features, the two configurations performed equally well.

In both cases, and in the two years, although the best dates varied, the selected time period was from early spring to late summer. In all cases, the best selected single feature matched a summer date. After the initial selection, no further features were removed from the subset.

4.2 Single-feature SFFS

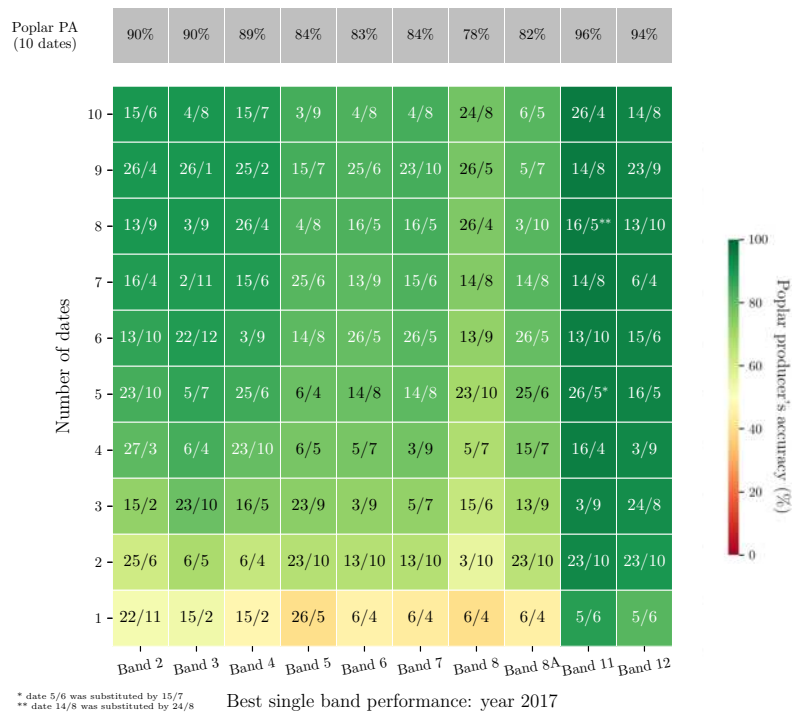
When we considered each spectral band separately and applied SFFS to all the dates in the 2017 (34 dates) and the 2018 (36 dates) time series, B11 and B12 from the SWIR spectral range achieved the best poplar PA values with the fewest features. Indeed, maximum performance was obtained with four features and reached 96% and 97% with B11 and 94% and 96% with B12 in 2017 and 2018 respectively. The performance was slightly better with B11 but was 2% lower than that achieved with the best combination of bands (i.e. in the multi-band configuration; section 4.1).

Beyond four features, the PA values no longer increased. Some dates were eliminated immediately after being selected and sometimes re-integrated in the following subsets. This was the case with B11 for which the date of August 14th was introduced in the subset of seven features and replaced by the date of August 24th in the following subset of eight features but selected again in the subset of nine features.

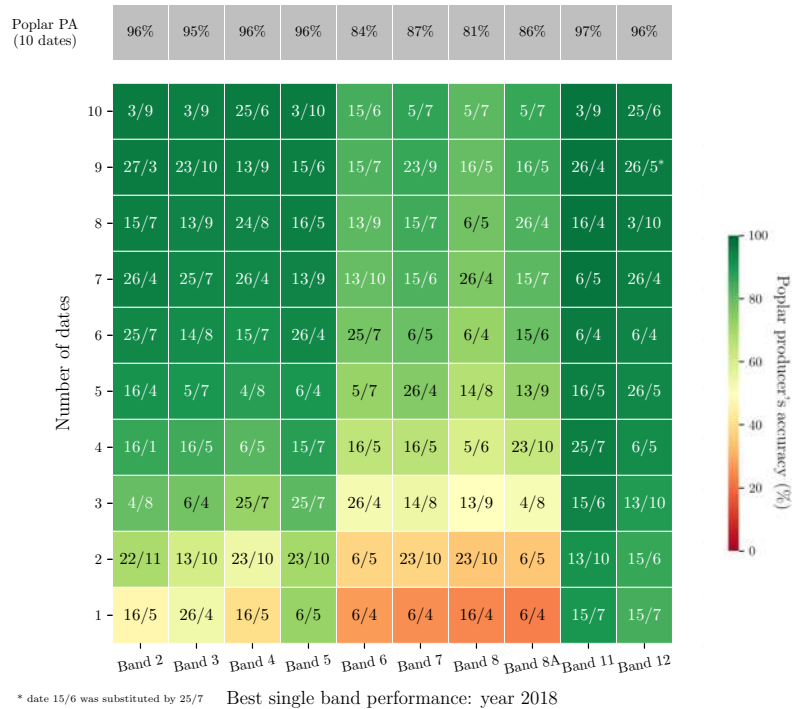
B5 and B6 of the red edge, ranked second in importance in the multi-band configuration after the SWIR bands, performed less well with the first dates but improved to reach their maximum value after six features. For instance, the poplar's PA of B5 was 40% with one date in 2017 but increased with additional dates and levelled off at 84%.

The same trend was observed for the remaining bands of the NIR (B7, B8 and B8A) and visible (B2, B3 and B4) domains whose initially low PA values increased with the addition of dates to reach their maximum performance, generally below that of the SWIR and red edge bands.

In the same SFFS configuration but using the 15 spectral indices instead of spectral bands, two stood out in both 2017 and 2018: PI_2 and SIWSI. The best poplar PA values were achieved



(a) Year 2017



(b) Year 2018

Figure 6.7 – Average poplar PA as a function of the number of features selected with SFFS in the single-band configuration. In (a) Single-band SFFS results in 2017 and in (b) Single-band SFFS results in 2018. The y-axis represents the number of features from 1 to a maximum of 10 for each individual spectral band on the x-axis. The colours on the red-yellow-green scheme palette reflect the poplar PA values. The selected dates are given inside the cells.

from three features, scoring 96% and 98% with PI_2 and 96% and 97% with SIWSI in the two respective years. Other indices also produced similar performances in the two years, including PI_1 and MSI. Although PI_1 performed best in 2017 with only one date (poplar PA = 91%), PI_2 slightly outperformed it with three dates to achieve the highest poplar PA in this single-index configuration. In 2018, PI_2 always performed the best regardless of the number of dates.

Like in the multi-feature configuration, the selection period ranged mainly between early spring and late summer. For the best features, the same summer dates were always identified first, namely the 5th of June in 2017 and the 15th of July in 2018.

4.3 Predictive performances using another independent dataset

In the previous results, the predictive performance used to select features was based on a 5-fold CV applied on 50% of the reference dataset, i.e. the training set (see section 3.2). Thus, the performances were computed on five independent subsets (named *validation* sets) by cross-validation. Here, performance values are given using the remaining 50% of references, i.e. the test set. The best results are summarised in Table 6.3 for each SFFS configuration and for both 2017 and 2018. The *test* score represents the poplar PA using the test set. The *validation* score represents the poplar PA based on cross-validation for comparison (results of sections 4.1 and 4.2). The results of both test and validation scores are based on the ten best subsets of features identified using SFFS.

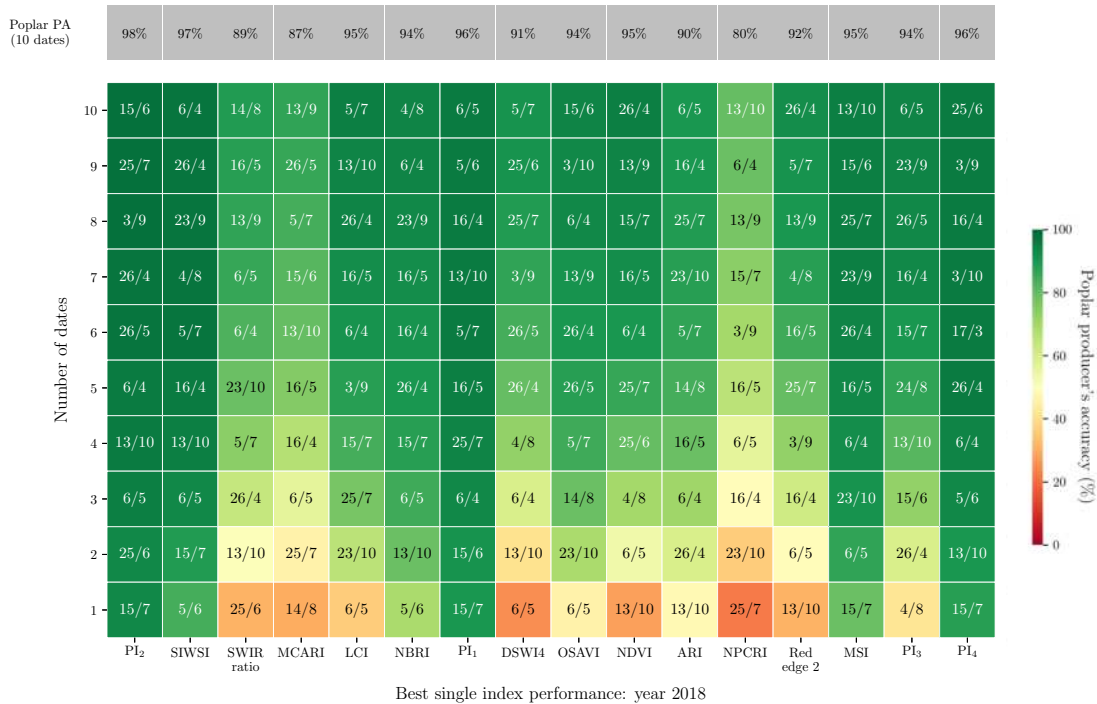
We observed consistent results with the test set compared to validation scores. The performance gap between test and validation scores ranged from 3% to 27% depending on the year and the SFFS configuration. Using PI_2 in the single-feature configuration, the difference was 3% in 2017 and 4% in 2018 whereas using B5, it reached 27% and 16% respectively, in 2017 and 2018.

Considering only the test set, the best performance in 2017 was obtained with PI_2 in the single-feature SFFS configuration with a poplar PA of 92%. In the multi-feature configuration, the combination of indices obtained the best PA (94%). According to the McNemar statistical test, the predictive performances differed significantly (p-values < 0.001; see Appendix B.4) in the single and multi-index configurations, indicating that the combination of indices produced the best result in 2017. But the combination of indices or bands (multi-feature configurations) produced equivalent performances (p-value = 0.395).

For the year 2018, PI_2 still produced the best performance of the single-feature configurations. The combination of bands ensured the highest poplar PA in the multi-feature configuration. In both cases, the poplar PA was 95% with no significant difference (p-value = 0.523). The results of the other single or multi-index configurations were less good.



(a) Year 2017



(b) Year 2018

Figure 6.8 – Average poplar PA as a function of the number of features selected with SFFS in the single-index configuration. In (a) Single-index SFFS results in 2017 and in (b) Single-index SFFS results in 2018. The y-axis represents the number of features from 1 to a maximum of 10 for each individual spectral index of the x-axis. The colours on the red-yellow-green scheme palette reflect the poplar's PA values. The selected dates are given inside the cells.

Table 6.3 – Poplar PA scores obtained using the 10 best feature subsets based on SFFS in the single and multi-feature configurations. The test score gives the poplar PA after prediction based on the test set(50% of reference data). The validation score is based on a 5-fold CV applied on the training set(50% of reference data used for feature selection). The highest test scores are highlighted in bold.

| Year | Poplar PA scores | Single-feature SFFS | | | | | Multi-feature SFFS | |
|------|-------------------------------|---------------------|-----|-----|-----------------|-------|--------------------|---------------|
| | | Bands | | | Indices | | Multi-bands | Multi-indices |
| | | B5 | B11 | B12 | PI ₂ | SIWSI | | |
| 2017 | test score | 57% | 87% | 89% | 92% | 89% | 93% | 94% |
| | validation score ¹ | 84% | 96% | 94% | 96% | 96% | 98% | 98% |
| 2018 | test score | 80% | 93% | 90% | 95% | 87% | 95% | 94% |
| | validation score ¹ | 96% | 97% | 96% | 98% | 97% | 99% | 99% |

¹ see the best results using 10 features in sections 4.1 and 4.2.

4.4 National map of poplar plantations using the PI₂ index

The first map of poplar plantations at the French national scale was produced for the year 2018 using the PI₂ index in a single-feature configuration. This simple configuration ensures optimal performance with stable results regardless of the year. A confidence map associated with the classification was also produced with a confidence level ranging from 0 to 100% for the class assigned to each pixel. The closer the confidence value to 100, the more confidence one can have in the algorithm for the predicted class.

Some extracts are illustrated in Figures 6.9-6.11. Mature plantations were generally well detected with a high confidence level $\geq 75\%$ (Figure 6.9). In contrast, young stands with limited canopy covers were either not mapped or detected with a confidence level of less than 50%. In some stands for which the year of plantation was known, as expected, plantations less than three years old were not detected. However, in the case of plantations between four and six years old, the age limit for detection varied. Indeed, in Figure 6.9c, four-year-old plantations were detected with a confidence level of between 50% and 90% whereas in Figure 6.9f, six-year-old plantations were not systematically detected.

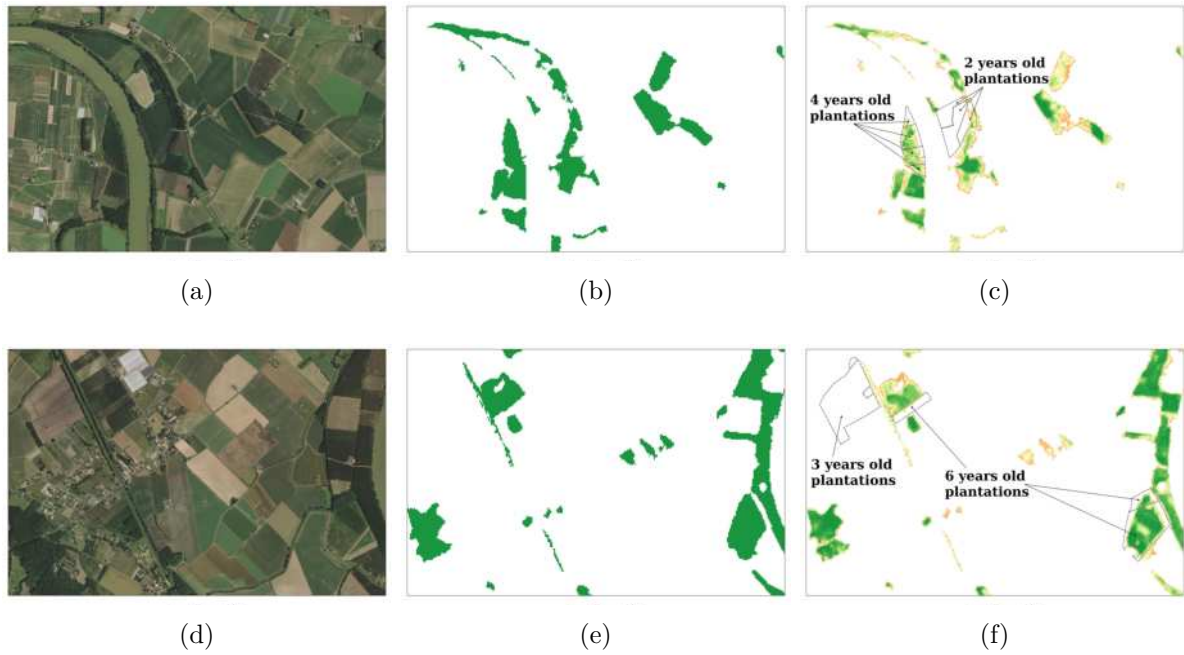


Figure 6.9 – Map extracts located in the Lot-et-Garonne department in tile 30TYQ. (a) and (d) are the 2018 orthophotos, (b) and (e) show the classification results and (c) and (f) are the confidence maps associated with the red-yellow-green scheme palette (the highest confidence rates are in green).

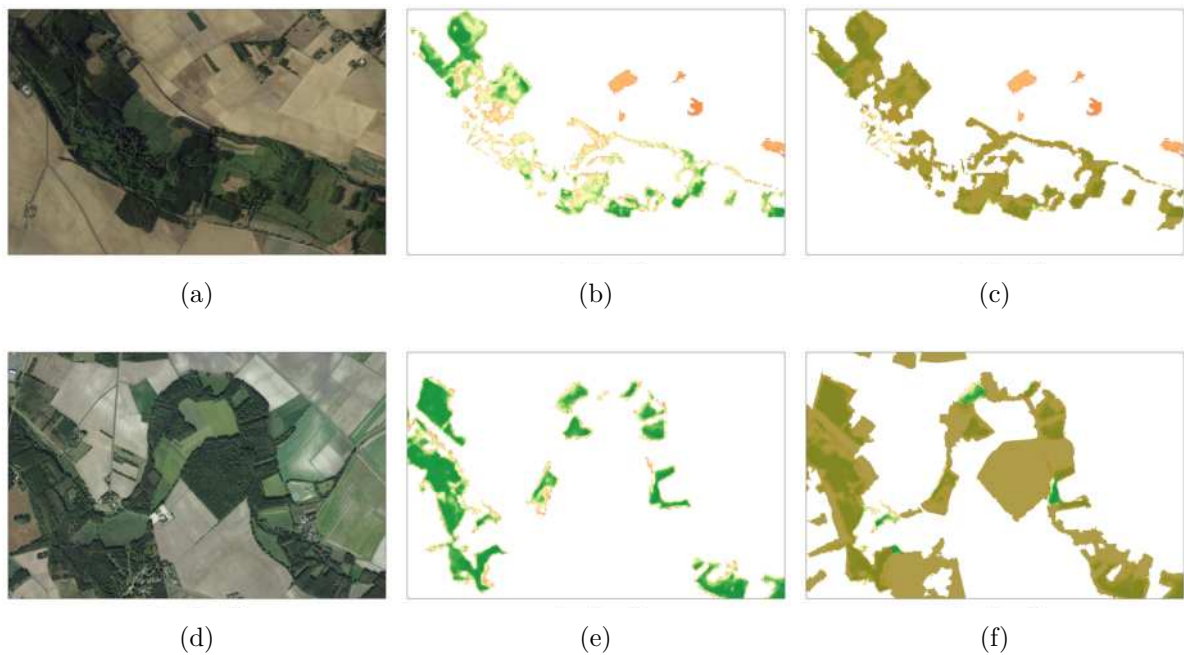


Figure 6.10 – Map extracts located in the Indre department in tile 31TCM. (a) and (d) are the 2018 orthophotos, (b) and (e) are the confidence maps with a red-yellow-green scheme palette for the classified poplars (the highest confidence rates in green), (c) and (f) show the overlay of the forest mask derived from the national forest database (BDForêt@IGN, v.2).

Over-detection of poplar plantations in agricultural fields was also observed (Figure 6.10a). In general, these cases were associated with low confidence values ($\leq 25\%$) (Figure 6.10b) and can be excluded using a forest/non-forest mask (Figure 6.10c). Nevertheless, because

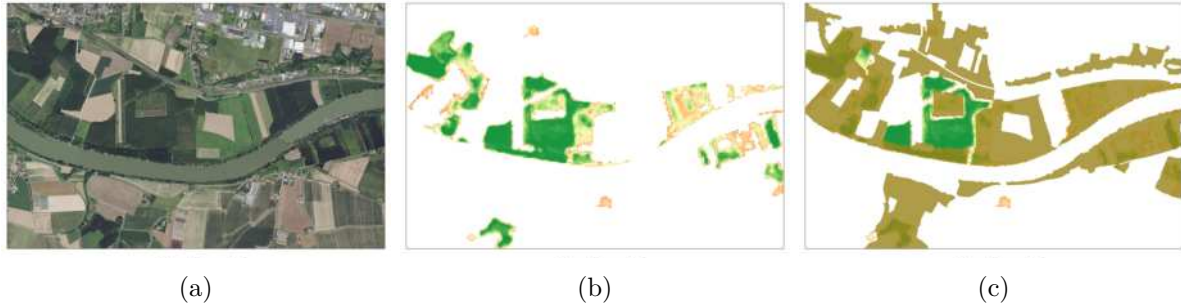


Figure 6.11 – Map extracts located in the Aquitaine department in tile 30TYQ. (a) is the 2018 ortho-photo, (b) the confidence map with a red-yellow-green scheme palette for the classified poplars (the highest confidence rates are in green) and (c) shows the overlay of the forest mask derived from the national forest database (BDForêt@IGN, v.2) on the confidence map.

of possible imperfections in the forest/non forest mask, some poplar plantations with high confidence values were also excluded (see Figures 6.10f and 6.11c).

Figure 6.12 shows the national map of poplar plantations with their prediction confidence level. A first estimate of the area under poplar based on pixel counting was produced for each municipality. Poplar forested areas are mainly distributed in the north, central and southwest France, as expected. Non-forest areas were masked using the 10-m Copernicus HRL 2018 forest layer. This mask helped to reduce over-detection in agricultural fields while retaining identified poplar stands.

5. Discussion

In this study, two years of 10-day resampled S2 time series were used to identify the optimal set of features and period needed to distinguish poplar plantations at broad scale. Based on a single-feature configuration using SFFS, we investigated the ability of specific wavelength regions and spectral indices with their respective dates to classify poplars. We also assessed the interest of combining multiple bands or indices using a multi-feature SFFS configuration. A new poplar detection index based on a SWIR and red edge S2 bands was defined and used to map plantations at the French countrywide scale. To our knowledge, this is the first attempt to analyse the most discriminating spectral regions of poplar plantations with S2 time series and to provide a S2-based poplar index to map poplar plantations in different management contexts over large areas.

5.1 Single-feature *versus* multi-feature: less is more

We observed no difference in accuracy between combinations of spectral bands or indices in the multi-feature configuration. This was true for both years even when only two features were selected. The results we obtained using single and multi-feature configurations with the ten best feature subsets – whether by cross-validation or on another test set – were very close

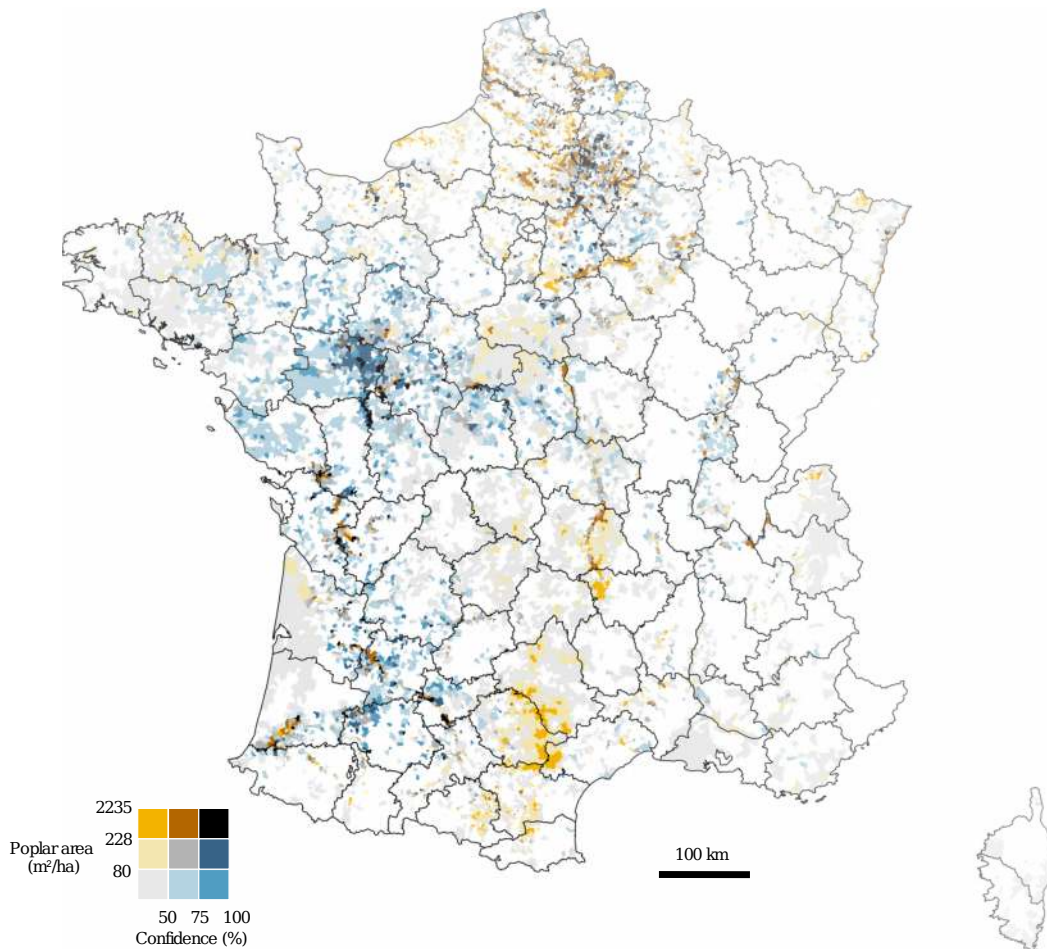


Figure 6.12 – Bivariate choropleth map of France showing the density of poplar plantations in each municipality with the median prediction confidence values for the year 2018. The map was produced from the PI_2 index using each image of the 2018 S2 time series (10-day frequency after resampling). Municipalities with no plantations or with a planted area of less than 5 ha are in white. The black lines represent the administrative boundaries of the 96 French departments. Non-forest areas were masked using the 10-m Copernicus HRL 2018 forest layer.

($\Delta_{PA} \leq 2\%$). The single PI_2 index at 10 dates provides almost the same performances as a combination of six different spectral bands at 8-10 dates or a combination of 8-9 spectral indices at nine dates. This suggests that a simple approach based on a specific multi-temporal single feature is adapted to classify poplar plantations accurately. The temporal information compensates for the spectral information provided by the combination of multiple bands or indices. The increase in precision with an increase in the number of features obtained with the PI_2 index quickly reached a plateau beyond four features (i.e. dates) spread out between May and October. Thus, a very parsimonious approach with only few dates can be used to distinguish poplar plantations with the highest performance. This makes processing highly efficient in terms of computational complexity and load. This strategy for mapping poplars over large areas complements the previously defined strategy based on active learning (Hamrouni et al., 2021). The advantage of the latter is that it requires fewer reference samples

but its implementation involves a more complex learning process.

5.2 S2 spectral bands in the SWIR and red edge domains are required to identify poplar

The results of variable selection with SFFS showed the interest of the SWIR spectral region (B11 and B12) to distinguish poplar plantations from other deciduous species. The results also highlighted the secondary importance of the red edge domain and particularly B5. Interestingly, these findings are in line with those reported by Viinikka et al. (2020), who used airborne hyperspectral data with 460 spectral bands covering the visible, NIR and SWIR ranges to discriminate European Aspen from three other species in southern Finland. These authors found that the most discriminative bands of aspen trees were located in the red edge ($\rho_{724-727\text{ nm}}$) and SWIR ($\rho_{1520-1564\text{ nm}}$ and $\rho_{1684-1706\text{ nm}}$) ranges that roughly match B5 and B11 in Sentinel-2.

Theoretically, the reflectance in the SWIR range is predominantly affected by leaf water content (Gao, 1996). Changes in water content can be observed with the SWIR bands located near the water absorption features at approximately 970 nm, 1200 nm, 1450 nm and 1950 nm wavelengths of the spectrum (Curran, 1989) and most notably around 1450 nm and 1950 nm (Clevers et al., 2008) which roughly coincide with Sentinel-2 bands 11 and 12. A number of remote sensing studies showed that SWIR bands reflectance decreased with an increase in leaf water content (Tucker et al., 1986; Gao, 1996). For poplar, Zhang et al. (1999) analysed seasonal (June to September) trends in water consumption of trees in a temperate climate and showed that the increase in solar radiation was followed by an increase in evaporative demand. At the beginning of the summer season, the increase in transpiration was offset by an increase in the absorption of water available in the soil due to the typical capacity of poplar to exploit groundwater. Later in the season, soil water content decreased, and absorption reached a low plateau, leading to different adaptation mechanisms such as stomatal closure to control losses through transpiration. This behaviour is fully consistent with the temporal profiles of the two SWIR bands, where in early spring, SWIR reflectance decreases rapidly as a result of increased water consumption to reach a low plateau, which continues throughout the dry season (see Figure 6.4).

The red edge spectral region proved to be highly sensitive to the chlorophyll content of the vegetation (Gitelson et al., 1996; Dash and Curran, 2004) and has been used to estimate structural features such as leaf area index (LAI) (Dong et al., 2019; Sun et al., 2020) as well as to map forest species distribution (Zarco-Tejada and Miller, 1999; Hu et al., 2008). In deciduous tree species, photosynthetic activity increases during the spring growing season along with the concentration of chlorophyll resulting from foliage growth (Garbulsky et al.,

2011; Springer et al., 2017). Like with the SWIR bands, reflectance decreases at the start of the growing season reflecting an increase in photosynthetic activity and in chlorophyll concentration (Gitelson et al., 1996). The temporal signature of B5 reflects this trend with a decrease in reflectance from early April to mid September (see Appendix B.5). However, only a marginal difference can be observed between plantations of poplar and those of other deciduous species which may explain the minor importance of the red edge compared to that of SWIR for the discrimination of poplar plantations.

The PI_2 index combining both B11 and B12 and subtracting them from the red edge band B5 makes it possible to accentuate the difference between the reflectance of poplar and of other tree species. The lower the reflectances in the SWIR, the higher the value of PI_2 (Appendix B.5). This formulation was more competitive than the other variants of the Poplar index (PI_1 , PI_3 , PI_4) even if the results of PI_1 (based on B11 and B12 only) were very close to the results obtained in the single-index configuration. This was also true using B11 or B12 alone or the SIWSI spectral index (a normalized difference combining B11 and B8a) for which the difference in performance was significantly smaller (Appendix B.4). The PI index defined from S2 imagery could be adapted for other sensors and in particular, Landsat 8-9 by selecting bands 6 (SWIR 1 at 157-165 nm) and 7 (SWIR 2 at 211-229 nm).

5.3 The national map of poplar plantations requires validation in the field

A first version of the national map was produced using the PI_2 index. As the feature selection analysis revealed, the map could have been produced using only a few dates between May and October. However, because of cloudy acquisitions that vary from one region to another, the possible influence of residual noisy pixels (undetected clouds), and the existence of some evergreen deciduous species (e.g. eucalyptus plantations), we used the full year time series of PI_2 to generate the national map (36 dates). After checking, we found that classification performances were not affected by the addition of all the dates (no Hughes phenomenon).

This map now needs to be validated in the field by forest partners to detect specific confusion with other species (missing from the reference set), the possible influence of the understorey vegetation in some regions where plantations were abandoned, and the detection limit related to the stage of development of the plantations. We assumed that no plantations less than three years old were mapped due to insufficient canopy cover and a possible effect of soil but, in practice, the minimum age for detection is more gradual because other factors have to be taken into consideration such as site conditions and the cultivar planted. Plantations can present different growth patterns with, for instance, a maximum growth rate in the first two years or a slow growth rate at the beginning and an increase in the growth rate later

in the cycle (Bergante et al., 2010; Sarsekova, 2015). We conducted a first analysis with a specific dataset of reference samples (85 stands; see section 2.2) for which the cultivars and the age of the plantations were known. We observed an increase in the confidence values of the detected plantations with age but this depended to a great extent on the cultivar concerned (Figure 6.13). Additional references are required to better define the detectability threshold of plantations. The absence of confusion with other short rotation coppice (SRC) plantations such as willow (*Salix*) which have higher water requirements should also be checked. A previous study revealed a clear distinction between poplar and willow but willow was growing in a natural context and not in SRC plantations (Karasiak et al., 2019).

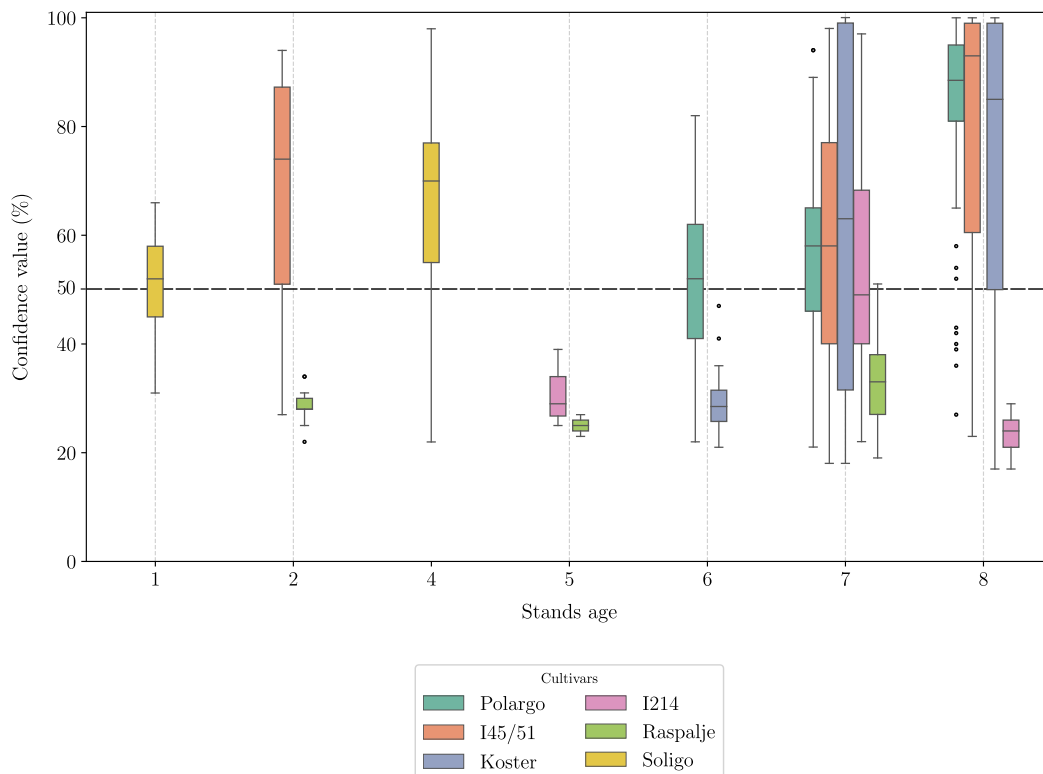


Figure 6.13 – Distribution of classification prediction confidence values as a function of the age of the detected plantations and of the cultivar planted. These values were extracted from 85 plots provided by the forest partners located in the *Lot-et-Garonne* department in southwest France (S2 31TCJ tile).

The national map identified the main poplar plantation sites in France but also a large number of municipalities with either a low density of poplar trees with a high level of confidence (in blue) or the opposite, a high density of poplar trees with a low level of confidence (in yellow). In the first case, in addition to the few plantations, this coincides with the presence of poplar in riparian areas. Some natural patches of poplar were detected along waterways. No precise evaluation was carried out to estimate the true ability to discriminate these poplars but it opens up possibilities for such areas of high conservation value (Villar and Forestier, 2017). Over detection was also observed in some places with low confidence values despite a

high density of poplars. Closer examination revealed that these cases often refer to confusions with coniferous stands (or agricultural fields). From the modelling point of view, these errors are not surprising since only references of deciduous species were used for training. Rather, they highlight the imperfection of the forest/non-forest mask used to only retain areas with deciduous species. The HRL Dominant Leaf Type 2018 was considered the best currently available candidate but its quality directly influenced our poplar plantation layer. In the future, the reference dataset could be enriched with conifer samples derived from the French National Forest Inventory spatial database to distinguish between coniferous and deciduous trees in the classification process. An alternative would be the adoption of a novelty detection approach to identify test data (unseen pixels) that differ significantly from the training set (Markou and Singh, 2003). This would make the method more independent of the existence of an accurate forest/non-forest mask.

6. Conclusions

The present study investigated the potential of S2 time series to discriminate poplar plantations over large areas including their diversity related to cultivars, development stages, management practices and climatic conditions.

Based on the results compared over two years, we draw the following conclusions:

- There is no added value to be obtained by combining multiple spectral bands or different spectral indices at different dates to classify poplar accurately. If optimal features are selected, a multi-temporal single-feature approach provides equivalent results.
- Only a few dates are required to identify poplar among deciduous species, the optimal period being from the middle of May to the beginning of October; no significant improvement is achieved beyond four or five acquisition dates but adding more dates can make the classification more robust to residual noise at national scale without being affected by the curse of dimensionality (because of the single-feature strategy).
- SWIR followed by red edge spectral regions are the most useful to differentiate poplars from other deciduous species. This reflects the sensitivity of poplar trees to water content throughout its phenological cycle. The best S2 spectral bands are B11, B12, B5 and B6. The best performances with stable results regardless of the year were obtained when some of these bands were combined through the PI_2 poplar index. Significant but limited differences were found with the PI_1 or SIWSI indices (which are the other competitive ones).
- Because the model was trained using reference samples of deciduous species only, the national map of poplar plantations strongly depends on the quality of the forest/non-forest layer used to mask the unfocused areas.



The short-term future outlook of this work is to produce a national map of poplar plantations for other years. Comparing the different years should enable better evaluation of the potential errors in the maps, and limit a possible year effect in some regions. Comparing multi-year production is also the first step towards monitoring poplar resources and detecting harvest. It should help overcome the challenges related to young plantations which were not reliably detected due to their open canopy but would become detectable over the years.

Another future objective is to make the approach less dependent on the availability and quality of a forest/non-forest mask. The detection of forest tree cover before distinguishing poplar is one possible option, based on S2, but also multitemporal Sentinel-1 data (Dostálová et al., 2021). This would ensure an up-to-date mask for each annual production of the poplar layer.

6.3 Bilan

Dans ce chapitre, nous avons déterminé le potentiel des régions spectrales du SWIR et du *Red edge* des données Sentinel-2 pour différencier les peupliers des autres essences de feuillus, en combinant trois bandes spectrales qui en sont dérivées, débouchant sur la proposition d'un nouvel indice spécifique au peuplier, le PI_2 . Le profil temporel du PI_2 reflète la sensibilité des peupleraies à la teneur en eau, avec une démarcation assez nette par rapport aux autres essences pendant la période printemps-été, un comportement qui se révèle stable sur les deux années étudiées.

Cette démarche est à la fois spécifique au peuplier, simple et facile à mettre en œuvre pour une production à l'échelle nationale. Elle implique une classification supervisée classique basée sur les séries temporelles du PI_2 et d'échantillons issus des trois tuiles de référence représentant la variabilité nationale.

Par ailleurs, comme elle ne fait pas appel à des algorithmes très complexes comme cela peut être le cas avec certaines approches d'adaptation de domaine, elle est parfaitement compatible avec la chaîne Infrastructure pour l'Occupation des sols par Traitement Automatique Incorporant les Orfeo Toolbox Applications (*iota*²) dédiée à la production cartographique opérationnelle à partir de gros volumes de données satellitaires et utilisée pour la génération de la carte nationale d'occupation des sols du CES OSO.

L'originalité de cette approche n'est certainement pas méthodologique, mais elle réside dans sa spécificité au peuplier en proposant un indice dédié, son potentiel de discrimination dans trois contextes popuicoles contrastés et sa stabilité temporelle entre années.

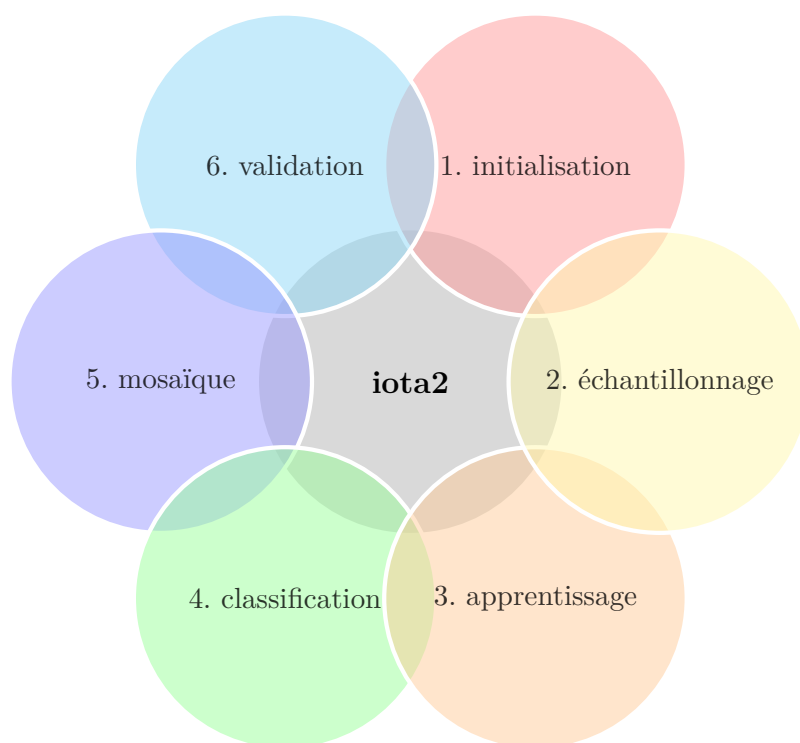
Pour toutes ces raisons, l'approche basée sur le PI_2 a été retenue pour la production opérationnelle à l'échelle nationale.

Chapitre 7

Production à l'échelle nationale

Sommaire

| | | |
|-------|--|-----|
| 7.1 | Projet TOSCA PARCELLE | 178 |
| 7.2 | Présentation de la chaîne de traitement <i>iota</i> ² | 178 |
| 7.3 | Présentation de l'infrastructure de calcul haute performance du CNES . . . | 180 |
| 7.4 | Utilisation de <i>iota</i> ² pour la production de la carte nationale des peupleraies sur le cluster HPC du CNES | 181 |
| 7.4.1 | Données d'entrée | 182 |
| 7.4.2 | Configuration de la chaîne : étapes de traitement | 183 |
| 7.4.3 | Post-traitements | 186 |
| 7.4.4 | Diffusion et maintien de la production à long terme | 187 |



Dans les chapitres précédents, nous avons vu qu'au-delà de l'aspect méthodologique lié au transfert de connaissances sur de grandes étendues, la production de la carte des peupleraies à l'échelle nationale avec la dernière approche basée sur le PI_2 exige une grande capacité de stockage et de traitement du volume de données Sentinel-2 couvrant l'ensemble du territoire.

Tous les traitements effectués jusqu'à présent, notamment les pré-traitements, l'adaptation de domaine et la classification supervisée, ont été réalisés sur les trois tuiles de référence à l'aide de codes Python faisant appel à différentes bibliothèques détaillées dans chacun des chapitres concernés. Les données utilisées ont été traitées et stockées sur un serveur interne au laboratoire, ce qui convenait tout juste aux trois tuiles. Pour une production nationale, au regard des moyens nécessaires, il était nécessaire d'envisager une infrastructure de calcul plus adaptée. Il s'agit de la principale contrainte technique rencontrée et qui a été appréhendée à travers la mise à disposition de supports performants dans le cadre de collaborations avec d'autres organismes. Dans ce chapitre, nous présentons les moyens et outils mobilisés pour ce passage à l'échelle, aboutissant à la production d'une première carte des peupleraies en 2018 sur le territoire métropolitain présentée à la fin du Chapitre 6.

7.1 Projet TOSCA PARCELLE

Nous avons été impliqués dans le projet TOSCA Production cARtographique grande éChELLE par tÉLédétEction (PARCELLE) porté par le CESBIO et mutualisant les efforts de huit laboratoires de recherche (CESBIO, DYNAFOR, CIRAD-UMR TETIS et Eco&Sols, INRAE-UMR TETIS, LaSTIG, IRISA, ESPACE-DEV et Geosciences Environnement Toulouse) afin de développer l'utilisation de la chaîne de traitement *iota*², conçue pour la cartographie sur de grandes étendues. Le projet vise d'une part à évaluer le potentiel de la chaîne de traitement sur des milieux variés (zones urbaines, forêts, milieux semi-naturels, milieux agricoles, etc.) avec des nomenclatures différentes. D'autre part, il s'agit d'intégrer dans la chaîne de nouvelles approches méthodologiques telles que l'apprentissage profond ou l'incorporation d'informations contextuelles, toujours dans un contexte opérationnel sur de grandes étendues.

Dans ce cadre, la chaîne *iota*² a été évaluée pour la cartographie nationale des peupleraies au titre de l'année 2018 et une production est également envisagée après la thèse pour les deux années 2019 et 2020.

7.2 Présentation de la chaîne de traitement *iota*²

La chaîne *iota*² a été développée par le CESBIO afin de répondre aux besoins de traitements de gros volumes de données satellitaires sur de grandes étendues. Il s'agit d'un outil

libre sous la licence publique générale GNU Affero v3.0¹, accessible gratuitement² et en perpétuel développement. La chaîne *iota*² est développée en Python et elle est composée d'un ensemble de processus ou *workflow* allant des pré-traitements jusqu'au calcul des indicateurs de performance à l'issue de la classification.

La chaîne est structurée en plusieurs tâches regroupées en étapes qui sont à leur tour incluses dans des groupes. Une tâche correspond à l'exécution d'un traitement à partir d'une donnée d'entrée et dont la sortie peut être utilisée dans une tâche suivante, comme par exemple l'apprentissage d'un modèle de classification ou le calcul d'une matrice de confusion. L'étape gère plusieurs tâches qui peuvent être exécutées de manière séquentielle ou parallèle. L'avantage de la structuration en étapes est de pouvoir garder des données de sortie en mémoire sans avoir à écrire des résultats intermédiaires. D'autre part, comme les étapes sont indépendantes les unes des autres, si une erreur se produit à une étape donnée, les étapes terminées n'ont pas besoin d'être ré-exécutées et la chaîne peut être relancée à partir de l'étape inachevée.

Plusieurs étapes peuvent être rassemblées au sein d'entités abstraites nommées groupes, conçus à des fins de planification des exécutions et de configuration de la chaîne. La chaîne est conçue de façon à réaliser la plupart des traitements tuile par tuile et elle se base sur des bibliothèques Python telle que *dask* (Dask Development Team, 2016) permettant d'optimiser au maximum les ressources des plateformes de calcul avec une planification dynamique des tâches et un calcul parallèle.

La version actuelle de *iota*² comprend sept groupes (Figure 7.1) où chacun est constitué d'une série d'étapes. Par exemple, l'initialisation consiste en quatre étapes principales à savoir la vérification des données d'entrée et de leur cohérence, le pré-traitement des images incluant plusieurs tâches (ré-échantillonnage temporel, comblement des lacunes...), le calcul de l'emprise de chaque image, et la génération d'un masque de validité correspondant au nombre de pixels non nuageux. La chaîne supporte les données issues des quatre capteurs : Sentinel-1, Sentinel-2, Landsat-5 et Landsat-8 et l'ensemble de ces traitements ont été définis pour chaque type de données.

1. Licence libre avec la permission d'utilisation, de modification et de diffusion (licence *copyleft*) : <https://www.gnu.org/licenses/agpl-3.0.html>

2. http://osr-cesbio.ups-tlse.fr/oso/donneeswww_TheiaOSO/iota2_documentation/master/index.html



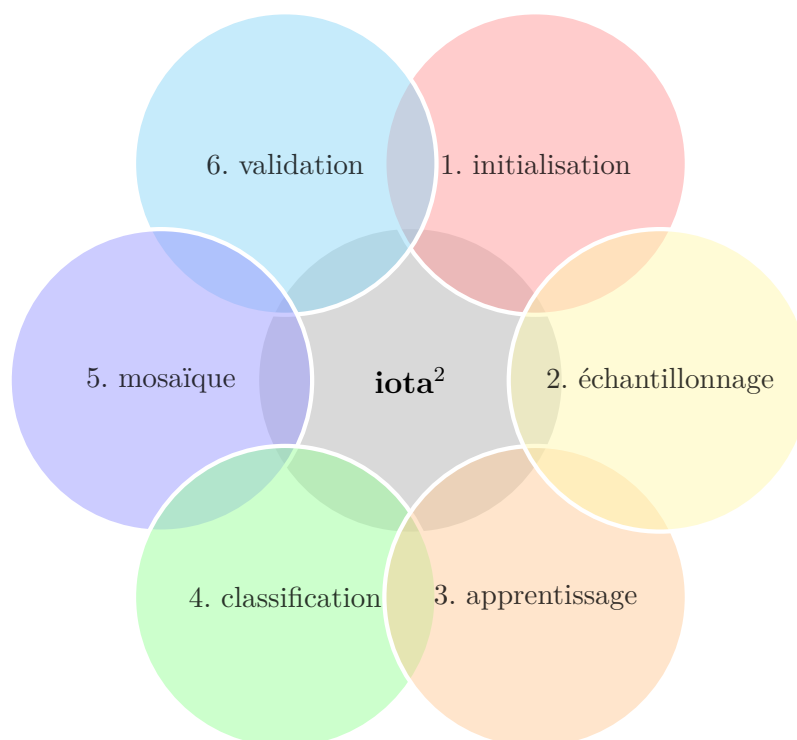


FIGURE 7.1 – Illustration des sept principaux groupes formant le flux de travail dans $iota^2$.

$iota^2$ est exécutée à partir de deux fichiers. Un premier fichier de configuration qui définit les chemins des données d'entrée et de sortie, les étapes à exécuter (toutes ou une sélection d'étapes), ainsi que les paramètres des différentes étapes au sein de chaque groupe. Un deuxième fichier spécifiant les ressources mémoire à allouer par groupe en fonction de l'infrastructure de calcul choisie (cluster ou cloud).

Cette flexibilité de configuration offre aux utilisateurs la possibilité de créer une chaîne personnalisée avec une sélection des différentes étapes à exécuter, la définition de nouvelles tâches, étapes, ou groupes ainsi que l'utilisation de données externes à la chaîne tout en respectant certaines normes et règles de nomenclature. Ainsi, $iota^2$ a été configurée pour la production nationale de la carte des peupleraies à partir de l'indice PI_2 issu des séries temporelles Sentinel-2 de 2018.

7.3 Présentation de l'infrastructure de calcul haute performance du CNES

Dans le cadre du projet PARCELLE, nous avons pu bénéficier des moyens de calcul à haute performance *High Performance Computing* (HPC) du CNES pour réaliser la cartographie nationale. Le cluster HPC offre une infrastructure à forte puissance de calcul utilisant un grand nombre de processeurs, une grande quantité de mémoire et un volume de stockage important.

Nous avons notamment eu accès à trois nœuds principaux de calcul interactif de type *Red Hat* (RH7) dont les caractéristiques techniques sont résumées dans le Tableau 7.1. Ils permettent d'effectuer des traitements lourds avec un calcul de manière parallèle.

TABLEAU 7.1 – Caractéristiques et ressources accessibles sur le cluster HPC (CNES, 2021).

| Nœud de calcul | Nb de CPUs | Mémoire/CPU | Mémoire totale | Nb nœuds |
|----------------|------------|-------------|----------------|----------|
| RH7-g2013 | 16 | 3,8 Go | 61 Go | 11 |
| RH7-g2016 | 24 | 5 Go | 120 Go | 272 |
| RH7-g2019 | 40 | 4,6 Go | 184 Go | 100 |

Deux espace disques ont été mis à disposition. Un premier espace de type *General Parallel File System* (GPFS) ayant une taille de 2 To par utilisateur qui était dédié au stockage des fichiers d'entrée et de sortie lors des calculs. Un deuxième espace GPFS de 50 To spécifique au projet PARCELLE doté d'une haute performance de lecture et d'écriture et qui a été consacré à l'écriture et au stockage des résultats de la production nationale. D'autres nœuds de calcul dédiés à la production peuvent être attribués sur demande, permettant une plus grande allocation des ressources mémoire et une priorité dans l'exécution des traitements.

Enfin, nous avons eu accès à l'espace de stockage des données Sentinel-2 au niveau 2A³ pré-traitées par le pôle Theia et préalablement téléchargées dans le cadre de la production de la carte nationale d'occupation des sols du CES OSO (Inglada et al., 2017).

La chaîne de traitement *iota*² était déjà installée sur le cluster du CNES et sa mise à jour est assurée par les collègues du CESBIO.

7.4 Utilisation de *iota*² pour la production de la carte nationale des peupleraies sur le cluster HPC du CNES

La chaîne *iota*² a été configurée de manière à apprendre un modèle de classification supervisée basé sur les résultats du Chapitre 6 et entraîné à partir d'échantillons de référence provenant des trois tuiles d'étude situées dans le Sud-Ouest (31TCJ), le Centre (30TYT) et le Nord-Est (31UEQ). Ce modèle est ensuite utilisé pour prédire les autres tuiles Sentinel-2 couvrant l'ensemble de la France. Les trois tuiles sont considérées comme représentatives de la variabilité des peupleraies sur l'ensemble du territoire (Chapitre 2, Section 2.1).

3. Pour rappel, il s'agit des images orthorectifiées et corrigées des effets atmosphériques et des effets de pente (Section 2.2.2)

7.4.1 Données d'entrée

La France est couverte par 90 tuiles Sentinel-2 acquises selon le système MGRS. Les données sont ainsi fournies avec trois projections UTM selon les trois fuseaux qui traversent le territoire métropolitain : UTM 30, UTM 31 et UTM 32 (Figure 7.2).

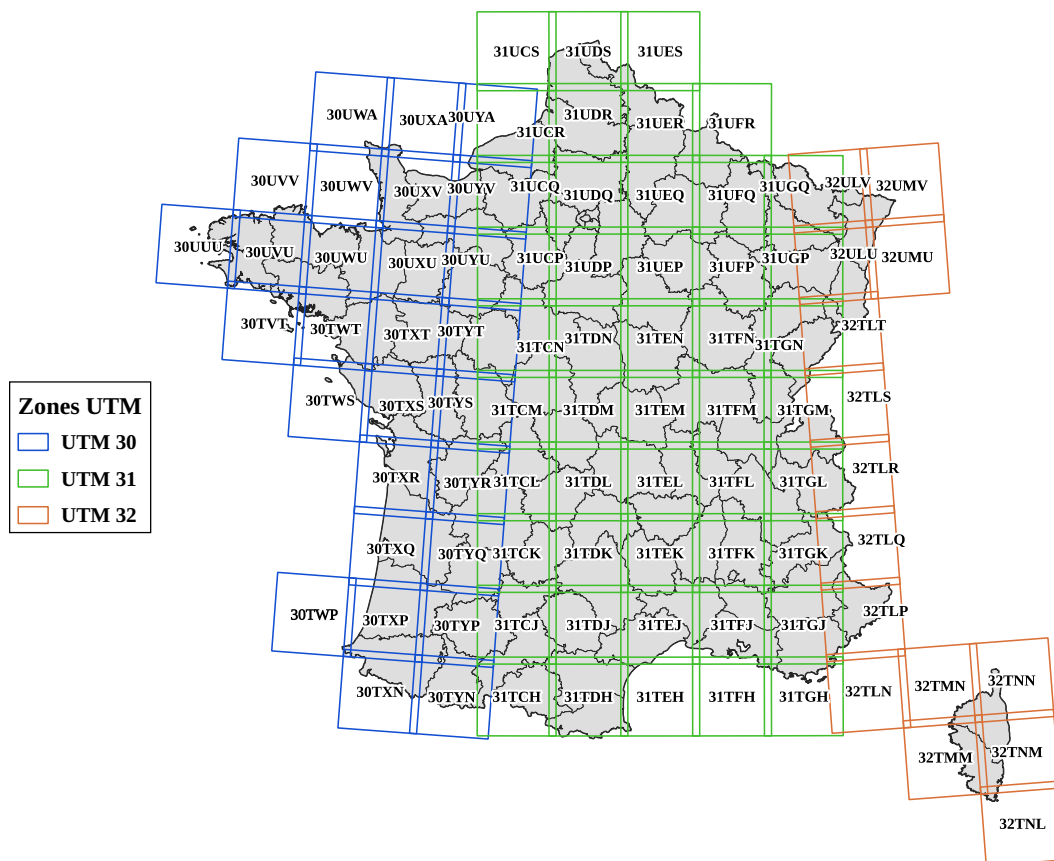


FIGURE 7.2 – Emprises des 90 tuiles Sentinel-2 couvrant l'ensemble de la France où chacune est identifiée par un nom de cinq caractères selon le système MGRS. Les couleurs représentent les trois zones UTM.

Toutes les acquisitions Sentinel-2 de 2018 sur l'ensemble de la France ont déjà été mises à disposition sur le cluster. Il s'agit de la première année d'acquisitions complète avec les deux satellites jumeaux Sentinel-2A et Sentinel-2B selon un cycle de cinq jours.

Sur les 90 tuiles, les données représentent un volume total d'environ 17 To incluant uniquement les produits FRE pour toutes les dates et pour toutes les emprises par tuile (deux à trois emprises par tuile dues à des orbites relatives différentes, exemple Figure 7.3) et comprenant les trois masques nécessaires aux pré-traitements à savoir : les masques de valeurs nulles, de nuages et de leurs ombres ainsi que le masque des pixels saturés (Section 2.2.2), tous à 10 m de résolution spatiale. Le fichier des polygones de référence de 2018 sur les trois tuiles d'entraînement comportait 512 polygones de peuplier et 4890 entités d'autres essences composés de six classes de feuillus.

Bien que $iota^2$ permette d'entraîner des modèles de classification selon une stratification spatiale comme par exemple des régions éco-climatiques (donc un modèle par région), cette approche n'a pas été retenue. Nous avons fait l'hypothèse, en collaboration avec les partenaires forestiers, que les trois tuiles couvraient toute la variabilité des situations de peupleraies en France incluant des stades de développement différents, une diversité de cultivars et de pratiques de gestion. Ainsi, un seul modèle de classification incluant les références des trois tuiles a été généré pour toute la France.

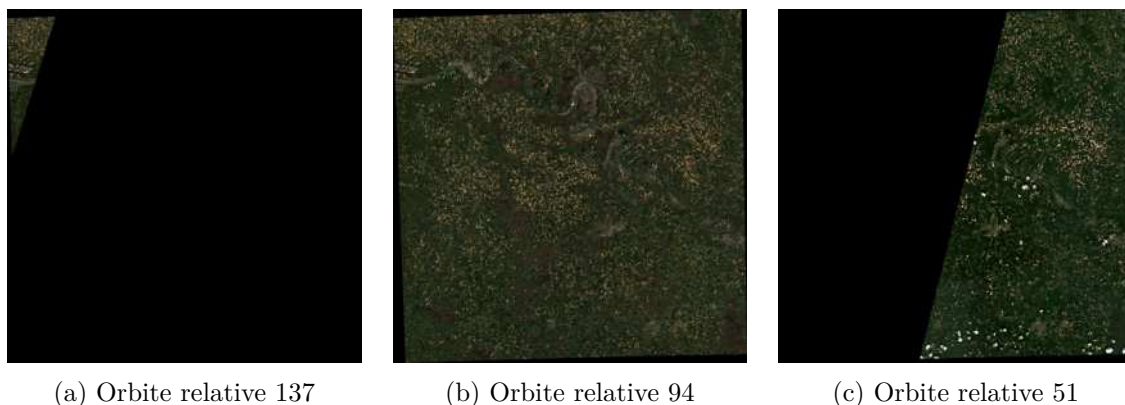


FIGURE 7.3 – Emprise de tuile 31UCQ selon trois orbites relatives.

7.4.2 Configuration de la chaîne : étapes de traitement

Les traitements compris dans $iota^2$ pour produire la carte nationale des peupleraies sont les mêmes que ceux appliqués dans la procédure de classification supervisée présentée dans la Section 3.1.2. Il s'agit de six étapes principales :

1. **La préparation des séries temporelles sur les 90 tuiles** : compte tenu de l'écart entre le nombre d'observations disponibles dû soit à une multiplication des acquisitions grâce à des recouvrements entre orbites adjacentes soit, au contraire, à l'absence de nombreuses acquisitions dans des régions plus nuageuses, chaque tuile a été pré-traitée (Chapitre 2, Section 2.2.3) en (1) corrigeant les pixels nuageux (*gap-filling*), (2) appliquant un ré-échantillonnage temporel homogène selon un intervalle de dix jours pour obtenir une fréquence d'images identique partout sur le territoire et (3) appliquant un ré-échantillonnage spatial pour obtenir toutes les bandes spectrales à 10 m de résolution spatiale. Le nombre d'acquisitions disponibles par tuile est récapitulé en Annexe C.1, Tableau C.1. Les dates ré-échantillonnées avec un pas commun de dix jours sont au nombre de 36 et elles sont fournies en Annexe B.1, Tableau B.1. Lors de cette étape, nous avons aussi calculé pour chaque tuile un masque de validité. Il recense, par pixel, le nombre des acquisitions non-nuageuses sur l'ensemble de l'année (Figure 2.7).
2. **La création des jeux de référence d'entraînement et de test** : $iota^2$ permet de

séparer les polygones de référence en 50% pour l'entraînement et 50% pour le test. Les pixels sont ensuite extraits avec le même nombre par classe qui est fixé de manière à ce que la plus petite classe soit entièrement échantillonnée. Bien que le nombre d'échantillons par classe a pu être fixé, *iota*² ne permet pas de contrôler celui des échantillons test. Le déséquilibre entre les classes a impacté l'évaluation des performances avec des métriques dépendantes de la prévalence, notamment l'OA ou le F-score. Cette séparation a seulement été appliquée au départ afin d'évaluer les performances, mais pour la production nationale toutes les données de référence disponibles ont été utilisées. L'étape d'élimination des valeurs aberrantes fondée sur l'EI (Chapitre 2, Section 2.4.2) n'a pas pu être intégrée car la chaîne ne le permettait pas encore lors de son utilisation. La production a ainsi été réalisée sans appliquer cette étape. À l'avenir, nous pourrions en tenir compte en faisant appel à notre propre fonction Python *outlier_removal.py* développée à cet égard.

3. **L'extraction de variables** : dans son fonctionnement par défaut, *iota*² considère toutes les bandes spectrales à toutes les dates dans la procédure d'extraction des variables d'entrée pour la classification supervisée. Cette étape a été modifiée à travers la définition d'une nouvelle fonction permettant de calculer l'indice PI_2 . Le paramétrage de la chaîne a été modifié pour ne prendre en compte que les séries temporelles du PI_2 . Chaque pixel est ainsi caractérisé par 36 valeurs de PI_2 relatives aux 36 dates des séries temporelles.
4. **L'apprentissage d'un modèle de classification** : *iota*² a été configurée pour utiliser RF comme algorithme de classification. Dans la version actuelle, l'optimisation des hyper-paramètres par validation croisée est incompatible avec le choix de variables personnalisées pour la classification (PI_2). Les valeurs ont donc été fixées en fonction des tests effectués en dehors de la chaîne à savoir, un nombre d'arbre de 100, une profondeur maximale des arbres de 25 et un nombre de variables pour séparer chaque nœud égal à 10 (Chapitre 3, Section 3.2.2).
5. **La prédiction sur les autres tuiles** : le modèle de classification résultant de l'étape précédente a été appliqué pour prédire l'ensemble des 90 tuiles. Comme précisé précédemment, le modèle utilisé pour la classification à l'échelle nationale est celui entraîné avec l'ensemble des données de référence disponibles. Pour chaque tuile, chaque classification résultante est accompagnée d'une carte de confiance associant à chaque pixel une valeur de 0 à 100% en fonction de la confiance du modèle vis-à-vis de la classe prédite.
6. **L'évaluation des performances** : Comme 100% des échantillons ont été utilisés pour l'apprentissage, cette étape bien que disponible, n'a pas été exploitée puisque ce travail d'évaluation a été mené avant d'utiliser la chaîne (Chapitre 6).

Malgré la grande quantité de mémoire allouée par les CNES, nous avons été contraints de



diviser la production en quatre zones où chacune a été traitée séparément. Le traitement pour la production automatique de la carte des peupleraies sur l'ensemble du territoire français a duré une semaine.

Le résultat cartographique à l'issue des six étapes est au format raster, fourni par tuile individuelle ou selon une mosaïque de l'ensemble des tuiles considérées dans le calcul (Figure 7.4). Une carte de confiance des prédictions est également associée à chacune des tuiles.



FIGURE 7.4 – Mosaïque de la classification des 90 tuiles Sentinel-2 à la sortie de la chaîne de traitement iota^2 . Pour chaque pixel de chaque tuile, une valeur de classe lui est attribuée parmi les sept classes de feuillus considérées.

Bien qu'ils soient en cours de développement, les traitements de vectorisation actuellement disponibles dans iota^2 sont spécifiques à la carte d'occupation des sols du CES OSO et n'ont pas pu être adaptés pour générer la carte des peupleraies au format vectoriel.

Afin de faciliter la manipulation des produits raster générés par iota^2 , des post-traitements spécifiques ont été effectués en dehors de la chaîne en faisant appel à des codes Python.

7.4.3 Post-traitements

En dehors de iota^2 , trois post-traitements ont été appliqués sur les rasters de classification et de confiance :

7. **Application du masque forêt** : comme illustré dans la Figure 7.4, tous les pixels Sentinel-2 couvrant la France ont été prédits, y compris des pixels d'eau, de zones urbaines, etc. Afin d'éliminer les surfaces hors forêt, un masque a été appliqué. Il s'agit du produit HRL de l'année 2018 fourni par l'*European Environment Agency* (EEA). C'est un masque forêt/non-forêt généré à partir des images Sentinel-2 de la même année sur l'ensemble de l'Europe. Nous avons utilisé le masque intermédiaire des principaux types dominants (*Dominant Leaf Type* (DLT)) qui distingue les forêts de feuillus et de conifères avant leur traitement visant à exclure certaines surfaces conformément à la définition de la forêt de la FAO. Ce traitement était destiné à éliminer les parcelles de moins de 0,5 ha et celles qui sont à vocation agricole. Il ne doit pas être pris en compte dans notre cas puisque les peupleraies en font partie. Tous les rasters de classification et de confiance ont ainsi été masqués par le produit DLT.
8. **Découpage départemental** : pour faciliter leur utilisation par les techniciens forestiers régionaux, les rasters de classification et de confiance ont été découpés par les limites départementales. Pour chacun des 96 départements métropolitain, un raster de classification et de confiance recensant tous les pixels de forêt identifiés et leurs confiances associées ont ainsi été générés.
9. **Extraction des peupleraies** : sur les rasters de classification et de confiances masqués, seuls les pixels de peupliers identifiés ont été extraits. Ceci a donné lieu à des images binaires de classification : 1 pour les pixels de peuplier et 0 pour les pixels de non-peuplier. Les images de classification ont été converties au format binaire, réduisant ainsi leur taille (une taille maximale de 4 Mo pour les départements les plus populeux). De la même manière, les images de confiance ont été masquées par les pixels de peupliers identifiés dans les rasters de classification. Elles ont été ensuite converties au format entier (entier 16 bits) avec des valeurs de confiance allant de 0 à 100%. La taille totale des images de classification et de confiance sur tous les départements métropolitains est de 500 Mo. À terme, une conversion des rasters départementaux au format vectoriel rendrait le produit plus approprié aux usages spécifiques sur le terrain (importation des cartes sur tablette ou smartphone) par les partenaires forestiers.

Quelques extraits cartographiques des rasters de classification et de confiance obtenus à l'issue des post-traitements sont illustrés dans la Figure 7.5.

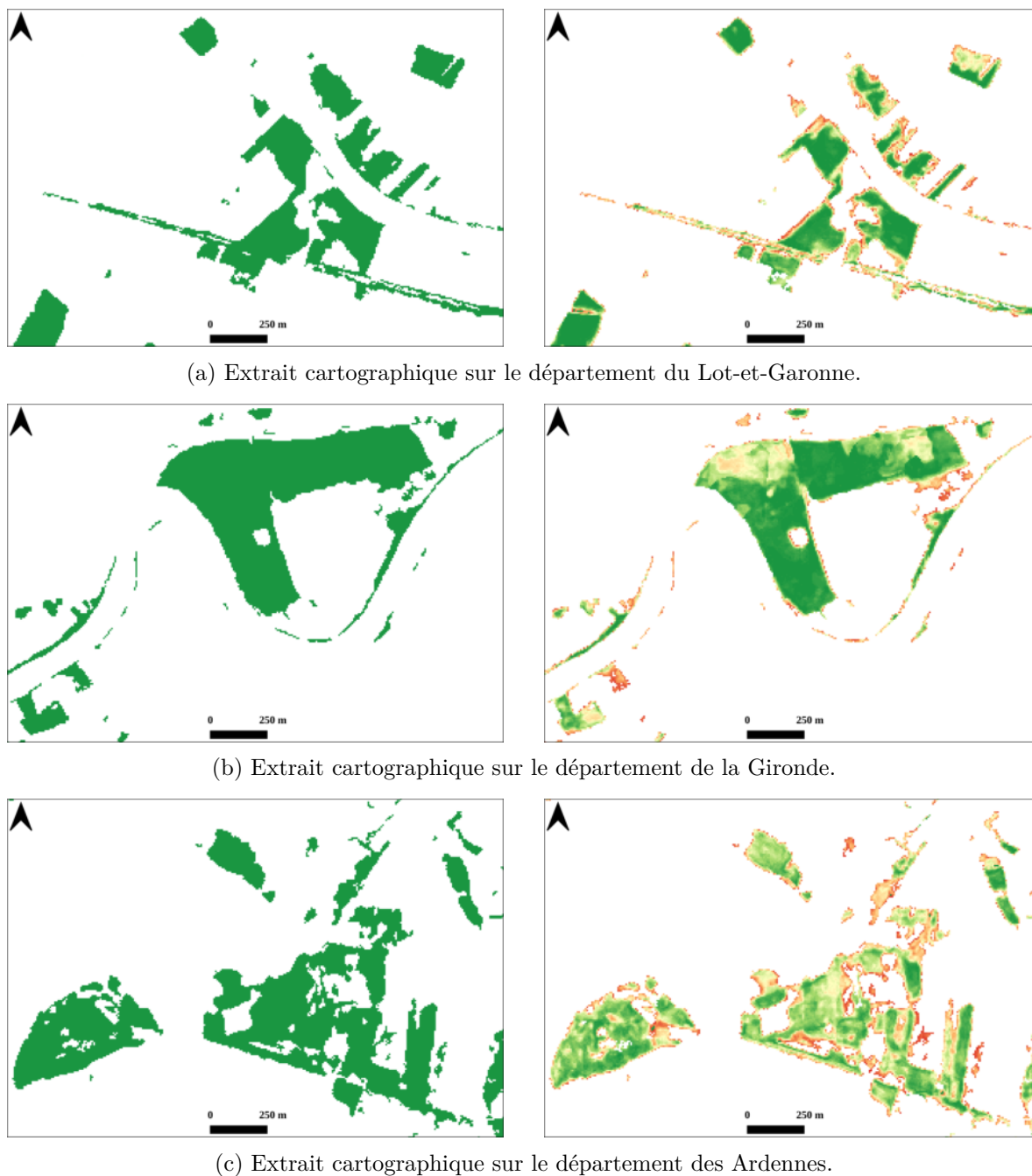


FIGURE 7.5 – Extraits cartographiques des classifications de peupleraies (gauche) et des confiances associées (droite) sur trois départements métropolitains. Sur les images de classification, les pixels de peuplier sont affichés en vert. Sur les images de confiances, les valeurs vont du rouge (faible confiance) au vert (forte confiance). Dans tous les extraits, les zones hors forêt ont été masquées par le masque DLT de 2018 fourni par l'ESA.

7.4.4 Diffusion et maintien de la production à long terme

Cette première version raster du produit peuplier pour le millésime 2018 a été transférée aux partenaires forestiers avec le découpage départemental. Dans l'attente de la validation sur le terrain et des premiers retours des utilisateurs, la production pour les millésimes 2019 et 2020 est prévue d'ici la fin de l'année. La chaîne étant configurée, ces productions nécessiteront

uniquement la mise à jour des références de peupleraies. Cela permettra de valider le produit sur trois années consécutives et de réduire les fausses détections due à l'effet d'une seule année.

Les spécifications du produit devront être être décrites afin de le préparer à une mise à disposition plus large. La diffusion est prévue par le pôle Theia, dans le cadre du CES OSO. Un protocole de mise à jour des références terrain sera fourni aux partenaires forestiers afin de permettre le maintien de la production, qui sera organisée à terme entre le CNP et le pôle Theia.

Cinquième partie

Conclusion

Chapitre 8

Discussion

Sommaire

| | | |
|-------|---|------------|
| 8.1 | Synthèse des principales contributions | 191 |
| 8.2 | Qualité des échantillons de référence | 194 |
| 8.2.1 | Imperfection des échantillons de référence | 194 |
| 8.2.2 | Impact de la présence des classes mixtes dans l'adaptation de domaine | 196 |
| 8.3 | Production cartographique | 197 |
| 8.3.1 | Limites pour la détection des jeunes plantations | 197 |
| 8.3.2 | Détection des peupliers naturels | 198 |
| 8.3.3 | Validation sur le terrain | 199 |
| 8.3.4 | Dépendance d'un masque forêt | 200 |
| 8.3.5 | Calcul des surfaces et post-traitements | 202 |
| 8.4 | Stabilité de la production nationale | 203 |

8.1 Synthèse des principales contributions

Dans le cadre de ces travaux de thèse, nous avons étudié le potentiel des données Sentinel-2 pour la reconnaissance des peupleraies à l'échelle de la France. L'enjeu de cette thèse est double. Il s'agit d'abord d'un enjeu méthodologique en proposant une démarche de classification à large échelle. Le second enjeu porte sur l'opérationnalité de l'approche proposée, avec de fortes attentes de la part de la filière peuplier pour une mise en œuvre simple assurant la production régulière d'une carte des peupleraies.

La principale contribution réside dans la mise en place d'une démarche opérationnelle fondée sur (i) l'adaptation de domaine avec un nombre limité d'échantillons de référence ou (ii) sans adaptation de domaine en considérant des échantillons multi-sites. Nous avons également proposé un nouvel indice spectral spécifique aux peupliers en analysant les bandes spectrales et les périodes de temps optimales. Enfin, nous avons produit la première carte nationale des peupleraies par télédétection.

Dans un premier temps, nous avons étudié la capacité des séries temporelles Sentinel-2 à identifier les peupleraies à des échelles locales dans trois contextes popuicoles distincts et contrastés en termes de pratiques sylvicoles et de conditions stationnelles (Chapitre 5, Section 3.2). Nous avons également évalué la qualité des données de référence requise pour la classification supervisée. Bien que nous ayons travaillé à des échelles locales, nous nous sommes placés dans un contexte de production au niveau national et nous avons considéré la BD Forêt comme source pour la collecte des références, de par sa disponibilité sur l'ensemble du territoire et de sa nomenclature homogène. Les résultats ont montré que la BD forêt seule n'est pas suffisante pour obtenir de bonnes performances de classification et que la mise à jour par photo-interprétation des échantillons de peupliers est nécessaire. Le gain en performance est de 6 à 12% en termes de PA de la classe peuplier. Dans le cadre de cette thèse, la photo-interprétation a pu être réalisée sur les trois sites et pour les deux années envisagées, mais dans une démarche opérationnelle de production annuelle, elle doit être répétée chaque année et sur chacune des 90 tuiles couvrant la France. Autrement, puisque les trois tuiles sont censées être représentatives de la variabilité nationale, il faut trouver un moyen de généraliser les connaissances acquises sur l'ensemble du territoire. Toutefois, les références photo-interprétées doivent inévitablement être vérifiées en raison du changement fréquent dans les peupleraies dû à leur exploitation. Afin de déterminer le potentiel des données Sentinel-2 à discriminer les peupleraies des feuillus dans un contexte de large échelle avec différentes contraintes de généralisation, nous avons mis en place trois stratégies en partant de la plus complexe à la plus simple, les deux premières étant deux techniques d'adaptation de domaine. Ce sont des techniques peu explorées en télédétection avec des séries temporelles d'images satellitaires

ayant à la fois une dimension spectrale et temporelle.

Dans la première (Chapitre 4), il s’agissait d’une adaptation de domaine par transport optimal, une approche d’adaptation des distributions des données qui est fondamentalement non supervisée. Pour rappel, l’hypothèse qui sous-tend le transport optimal stipule que si la distribution des données d’un site source connu (les distributions et les étiquettes sont connues) est transformée pour ressembler à celle d’un site cible moins connu (seules les distributions sont connues), un classifieur entraîné sur la source adaptée serait plus performant pour prédire la cible qu’un classifieur entraîné sur les données source non transformées. La version non supervisée du transport a conduit à des performances relativement limitées. Sa variante semi-supervisée a permis d’améliorer les résultats. En d’autres termes, il était nécessaire d’incorporer des étiquettes de la cible, ce qui nous amène à la question initiale du besoin d’échantillons étiquetés. Bien que la sensibilité au nombre d’échantillons cibles à incorporer ait été testée, leur sélection s’est faite de façon aléatoire, sans choix selon un critère défini. Par ailleurs, la complexité de la méthode du point de vue des classes à considérer, le temps nécessaire à l’optimisation des paramètres des algorithmes, mais surtout la complexité du transfert d’une telle approche à des partenaires non experts, ont fait du transport optimal un choix peu judicieux et relativement contraignant pour un déploiement à large échelle.

Afin d’envisager une approche plus opérationnelle tout en limitant le nombre de références pour apprendre, nous avons considéré l’adaptation de domaine par apprentissage actif. Dans cette approche, c’est le classifieur qui est adapté en tenant compte de nouveaux échantillons les plus informatifs et non directement la distribution des données (Chapitre 5). Il s’agit de modifier un classifieur entraîné sur la source en l’enrichissant, seulement si nécessaire, avec des échantillons de la cible en fonction de leur caractère informatif défini par un critère donné et ce, tout en restant performant sur la source. L’apprentissage actif est basé sur l’hypothèse qu’un algorithme d’apprentissage automatique peut atteindre de meilleures performances avec moins d’échantillons s’il est « autorisé » à choisir les données à partir desquelles il apprend. De ce fait, des échantillons étiquetés de la cible seront toujours requis mais en nombre limité et en sélectionnant les plus informatifs du point de vue de l’apprentissage. Cette approche a donné de bonnes performances de classification avec des valeurs de PA peuplier $\geq 90\%$ en utilisant moins de 50% d’échantillons étiquetés qu’avec un apprentissage automatique classique, et a abouti à un modèle global construit avec un minimum d’échantillons issus des trois sites d’étude et capable de reconnaître les peupleraies dans les trois contextes populaires. Malgré les performances obtenues, l’adaptation de domaine par apprentissage actif présente certaines contraintes d’opérationnalité à l’échelle nationale pour deux raisons principales. La première concerne le temps de calcul où l’algorithme doit déterminer en amont parmi tous les échantillons ceux qui seront étiquetés, un temps de calcul qui augmente avec le nombre de

variables liées aux dimensions spectrale et temporelle. En plus de la complexité d'intégration des algorithmes d'apprentissage actif dans les chaînes de traitement destinées à la production nationale, en pratique, cette approche doit être exécutée de manière itérative afin d'identifier les pixels que l'algorithme sélectionne, puis de procéder à leur étiquetage soit sur le terrain soit par photo-interprétation et de réapprendre ensuite un nouveau modèle. La procédure doit donc être répétée plusieurs fois (jusqu'à ce que les performances requises soient atteintes) ce qui la rend compliquée à mettre en œuvre dans un contexte opérationnel.

Les résultats de l'adaptation de domaine ont montré la nécessité de rajouter, même à minima, quelques échantillons de la cible pour rendre le modèle robuste à la large échelle. Par ailleurs, que ce soit à l'échelle locale ou avec les deux approches d'adaptation de domaine, toute la dimension temporelle et spectrale des données Sentinel-2 a été considérée sans chercher à identifier les variables les plus pertinentes pour les peupleraies. Cela nous a amené à la troisième approche proposée qui, contrairement aux deux approches précédentes visant à limiter le nombre d'échantillons, consiste à exploiter tous les échantillons disponibles sur les trois sites mais en identifiant des fenêtres spectrales et temporelles au sein de toutes les variables (Chapitre 6).

Dans cette approche, nous avons cherché à identifier les domaines spectraux les plus adaptés à la discrimination des peupleraies de même que les périodes de temps favorables dans l'année. Nous avons également vérifié la stabilité de ces variables spectrales et temporelles dans l'espace et dans le temps en vérifiant la pertinence sur les trois tuiles et sur deux années. D'une année à l'autre, les échantillons doivent être inévitablement mis à jour pour assurer une production cartographique régulière. La sélection des variables les plus pertinentes a montré l'intérêt de la région du moyen infrarouge (SWIR) et *Red edge* pour la discrimination des peupleraies, où trois bandes qui en sont dérivées ont été combinées pour proposer un nouvel indice, dénommé l'indice du peuplier ou *Poplar Index* (PI_2). La méthode d'identification des peupleraies avec le PI_2 a été déterminée et validée sur les deux années 2017 et 2018 des séries temporelles Sentinel-2. Par ailleurs, le déploiement en parallèle de la chaîne *iota*² conçue pour le traitement de gros volumes de données d'images satellitaires, dont Sentinel-2, a permis de lancer la production de la première carte des peupleraies sur l'année d'étude la plus récente, à savoir 2018.

Les trois stratégies proposées dans cette thèse permettent le transfert à large échelle mais avec un gradient de complexité différent. Les deux approches d'adaptation de domaine par transport optimal et avec apprentissage actif ont l'avantage d'exiger moins d'échantillons mais elles sont plus fastidieuses à mettre en œuvre et plus complexes pour un transfert à des partenaires non spécialistes. Si elle exige davantage d'échantillons de référence, la troisième approche avec le PI_2 s'est révélée plus facile à déployer à grande échelle.



Bien que seule la troisième approche ait fait l'objet d'une production cartographique à l'échelle nationale, toutes les approches ont permis de mettre en évidence différents enjeux, notamment en termes de qualité des données de références et de production cartographique. Les principaux enjeux seront abordés dans ce qui suit.

8.2 Qualité des échantillons de référence

8.2.1 Imperfection des échantillons de référence

Le « bruit » peut fortement impacter les résultats de classification et il est nécessaire d'adopter des stratégies pour en limiter l'impact. Il peut y avoir du bruit à la fois dans les images et dans les références. Dans les images, cela concerne surtout des nuages et des ombres non détectés. Dans les références, il peut s'agir d'un décalage entre l'étiquette attribuée et la réalité terrain selon la source de données utilisée.

Pour s'en prémunir, plusieurs approches ont été mises en œuvre : (i) une approche concernant les données et qui consiste à détecter les valeurs aberrantes (*outliers*) (en plus de l'utilisation des masques de nuage). Il s'agit d'une approche assez stricte, qui élimine le pixel dès qu'une seule valeur dans une bande donnée à une date donnée est détectée comme aberrante. Elle peut éventuellement être améliorée pour être moins exclusive et conserver le plus grand nombre de pixels possible. (ii) une deuxième qui concerne les références et qui consiste à vérifier les étiquettes des peupleraies par photo-interprétation.

En fonction de la méthode adoptée à l'échelle locale (par tuile) ou globale (avec transport optimal, apprentissage actif ou avec le PI_2), l'impact était différent.

La qualité des étiquettes des échantillons de référence a d'abord été évaluée dans le cadre de l'identification des peupleraies à des échelles locales au sein de chacune des trois tuiles de référence et seulement pour la classe des peupliers. La mise à jour des références peupliers par photo-interprétation s'est traduite par une augmentation de la performance de la classe de 9 à 17%. Lors de la photo-interprétation, la plupart des erreurs provenaient du décalage entre les années de référence de la BD Forêt et celles des acquisitions Sentinel-2, où des peupleraies étaient coupées et les parcelles sont devenues des friches ou des prairies (Figure 8.1b). Dans d'autres cas, les peupleraies coupées sont renouvelées et la parcelle peut correspondre à une très jeune plantation avec une faible couverture du sol (Figure 8.1c). Par ailleurs, d'autres erreurs étaient parfois liées à des imprécisions géométriques dans la délimitation des polygones, où l'ensemble du polygone est considéré comme une peupleraie alors qu'il renferme du bâti, du sol nu ou même de l'eau (Figure 8.1d). Pourtant tous les pixels qui y sont inclus seront étiquetés comme étant du peuplier tandis qu'ils représentent d'autres types d'occupation des sols. Dans d'autres cas un polygone peut englober plusieurs parcelles à différents stades de

développement (Figure 8.1a). Quelques exemples de polygones de peupleraies issus de la BD Forêt sont illustrés dans la Figure 8.1.



(a) Peupleraie avec différentes classes d'âge, des parcelles coupées et de très jeunes plantations



(b) Peupleraie abandonnée



(c) Très jeune plantation



(d) Vielle peupleraie coupée avec une délimitation grossière incluant une partie du cours d'eau

FIGURE 8.1 – Polygones de peupleraies dérivés de la BD Forêt illustrant différents exemples de bruit.

Ces erreurs n'ont été corrigées que pour la classe des peupleraies, notre classe d'intérêt pour laquelle une identification par photo-interprétation est possible. Pour les autres classes de feuillus, aucune correction n'a été appliquée et nous pouvons supposer la présence de bruit de la même manière avec un impact possible sur la reconnaissance, sans toutefois l'avoir évalué.

Ce bruit va se traduire dans l'espace des caractéristiques spectrales et temporelles qui décrivent l'échantillon. Deux pixels peuvent être attribués à la même classe alors que leur valeurs spectrales sont très différentes, ce qui peut induire le classifieur en erreur aussi bien dans la phase d'adaptation de domaine que dans la classification supervisée qui suit. Pour le transport optimal, cela peut entraîner des appariements entre des échantillons qui ne sont pas de la même classe et impacter par conséquent la fonction de transport. Cela a été observé dans le transport non supervisé sans information sur les étiquettes source et cible, où une grande confusion a été constatée entre les chênes et les robiniers, ce qui peut s'expliquer, au-delà du chevauchement des cycles phénologiques, par la présence de bruit dans les références issues de

la BD Forêt.

Dans le cas de l'apprentissage actif, si un échantillon est douteux (présente une faible valeur d'appartenance à la classe), il sera sélectionné par l'algorithme pour que son étiquette soit renseignée et utilisée dans le processus d'apprentissage. Cela s'est avéré être le cas pour la classe des robiniers et celles des forêts mixtes de feuillus où il y avait une contradiction entre l'étiquette de l'échantillon et l'apprentissage du modèle. Ce point concernant la présence de classes de mélange sera abordé plus en détails dans la section suivante.

Ce type de bruit concerne également les échantillons de test, ce qui peut avoir un impact sur l'évaluation des performances. D'autres types de bruit non liés aux étiquettes mais plutôt aux valeurs spectrales peuvent être introduits. Ils peuvent provenir notamment d'erreurs dans les corrections radiométriques ou lors de l'étape de comblement des lacunes avec la présence de nuages qui ne sont pas détectés dans les masques et ne sont donc pas corrigés.

L'étape de détection des valeurs aberrantes prévue à cet effet peut limiter en partie la présence de pixels très divergents de la distribution médiane de leur classe d'appartenance, mais plus le nombre d'échantillons bruités est important moins cette approche est efficace. Par ailleurs, le choix de l'algorithme de classification a une grande importance dans la mesure où certains classifieurs sont plus robustes que d'autres vis-à-vis de la présence de bruit. Plusieurs études ont comparé différents classifieurs en quantifiant l'influence des données mal étiquetées sur les performances de classification, et RF figurait souvent parmi les classifieurs les plus robustes (Folleco et al., 2008; Pelletier et al., 2016).

8.2.2 Impact de la présence des classes mixtes dans l'adaptation de domaine

La présence de classes de mélange dans les échantillons de référence peut être considérée comme un cas particulier de bruit évoqué dans le point précédent 8.2.1. Nous avons vu dans les chapitres 4 et 5 que le fait de considérer de telles références pouvait freiner l'amélioration des performances voire la pénaliser. En effet, dans la BD Forêt, les polygones sont définis comme étant des peuplements mixtes dès lors que l'espèce dominante occupe moins de 75% de leur surface. Par conséquent, ces polygones peuvent être composés de mélanges contenant des informations spectrales de plusieurs espèces dans un même pixel ou peuvent être un mélange de pixels purs.

La présence de ces pixels a posé deux problèmes. Premièrement, dans le jeu d'entraînement, avec l'adaptation de domaine par transport optimal dans le scénario 1, différentes paires ont été formées entre des pixels d'essences pures (ex. robinier, chêne) avec une classe mixte de feuillus (ex. les forêts de feuillus fermés).

Cela reste cohérent puisque des pixels de chêne ou de robinier sont inclus dans les classes

de mélange. Cependant, avant d'incorporer des étiquettes de la cible, les performances étaient impactées par la contradiction entre les classes d'appartenance. De même, avec l'apprentissage actif, les échantillons de robiniers et de feuillus fermés ont été considérés comme les plus incertains ce qui a perturbé l'algorithme et l'a conduit à ajouter de nombreux échantillons de ces deux classes, ralentissant ainsi l'évolution de la performance.

D'autre part, la présence des classes de mélange dans le jeu test servant à valider la classification avait un impact sur les performances de ces classes lors de l'apprentissage actif. Cette dérive a été renforcée en introduisant des échantillons de la cible étant donné que la composition des classes de mélange pouvait différer d'une tuile à l'autre.

Ainsi, nos résultats nous amène à recommander d'utiliser uniquement des échantillons de classes pures que ce soit à l'échelle locale ou avec les deux approches d'adaptation de domaine.

Plus généralement, dans l'adaptation de domaine pour la classification de la végétation, il est préférable de considérer des classes homogènes avec des échantillons de qualité afin que ces approches ne soient pas pénalisées, d'autant plus que même pour des pixels purs, il y a souvent un chevauchement entre le comportement spectral et temporel de certaines essences. En outre, les techniques d'adaptation de domaine gagneraient à une réduction de dimension, car plus la dimension de l'espace des caractéristiques est importante, plus l'évaluation entre la source et la cible est complexe.

8.3 Production cartographique

Dans le cadre de cette thèse, la première carte nationale des peupleraies par télédétection a été générée à partir de l'indice spectral PI_2 développé pour discriminer cette essence. Elle est fournie avec une carte de confiance qui traduit la fiabilité du modèle par rapport à la prédiction. Cette carte a été découpée par les emprises départementales afin de faciliter sa diffusion auprès des partenaires forestiers. Bien que la phase de validation de la carte sur le terrain avec un protocole bien défini ne soit pas encore effective, les premières interprétations ont mis en évidence certains enjeux qui seront abordés dans ce qui suit.

8.3.1 Limites pour la détection des jeunes plantations

Dans la production cartographique nationale, il était difficile d'estimer une limite précise sur l'âge de détection des jeunes plantations. En effet, les références ont été construites par photo-interprétation et nous n'avons généralement pas considéré les parcelles de moins de trois ans environ, ou du moins celles qui n'avaient pas une canopée bien développée, afin de limiter les risques de confusion avec d'autres types de plantations ainsi que l'influence du sol

nu. La vérification de l'âge des peuplements n'a donc pas été exhaustive, ce qui n'a pas permis d'identifier un seuil de détection précis.

Toutefois, grâce aux 85 parcelles fournies par les partenaires forestiers du département du Lot-et-Garonne avec des informations sur l'année de plantation et le type de cultivar, nous avons pu tirer des premières indications.

Suite à l'analyse du résultat cartographiques de ces parcelles terrain, nous avons conclu que le seuil de détection est très variable. En effet, l'âge de détection est en réalité fonction du taux de couverture de la canopée qui est à son tour lié au type de cultivar ainsi qu'aux conditions stationnelles telles que le type de sol, le climat ou la disponibilité des ressources en eau. En fonction de ces facteurs, les plantations peuvent présenter des profils de croissance différents, avec un taux de croissance maximal au cours des premières années qui suivent la plantation ou alors un rythme de croissance lent au départ, puis une accélération plus tard dans le cycle.

Comme nous ne disposons que de l'information sur le cultivar, nous n'avons pu confirmer que pour certains cultivars d'entre eux tels que le *I45* ou le *Soligo*. Ces deux cultivars sont généralement préconisés sur des sols riches avec un accès facile à la nappe phréatique, et pour lesquels l'âge de détection se situait entre 2 et 4 ans avec des valeurs de confiance associées supérieures à 50%. Tandis que dans le cas du *Polargo* qui requiert quasiment les mêmes conditions stationnelles que les deux précédents, la détection a commencé avec les parcelles de 6 ans. Bien que d'autres références terrain soient nécessaires pour confirmer ces déductions, on peut estimer que l'âge de détection se situe au plus tôt vers l'âge de 3 ans et au plus tard autour de 6 ans.

Ces déductions impliquent que l'estimation des surfaces à partir des résultats de la classification ne tiendra pas compte de ces jeunes plantations et sera par conséquent sous-estimée. Il s'agit d'une considération très importante, car l'estimation des surfaces représente une attente très forte de la part de la filière dont il faut tenir compte. Néanmoins, il convient de souligner qu'au fur et à mesure que les canopées de ces jeunes plantations se développeront, elles apparaîtront dans les millésimes suivants et pourront être prises en compte dans les estimations.

8.3.2 Détection des peupliers naturels

Les premières interprétations et les retours du terrain montrent que la carte produite avec le PI_2 identifie également des peupliers naturels soit comme des arbres isolés, soit alignés en bordure de cours d'eau (Figure 8.2). Cela souligne deux choses. D'une part, l'approche développée ne se réduit pas exclusivement à la détection des peupliers plantés mais à l'espèce peuplier en tant que telle, sans toutefois la contrôler. D'autre part, le modèle global généré

à l'échelle nationale comporte une certaine variabilité qui doit couvrir celle des peupliers naturels.

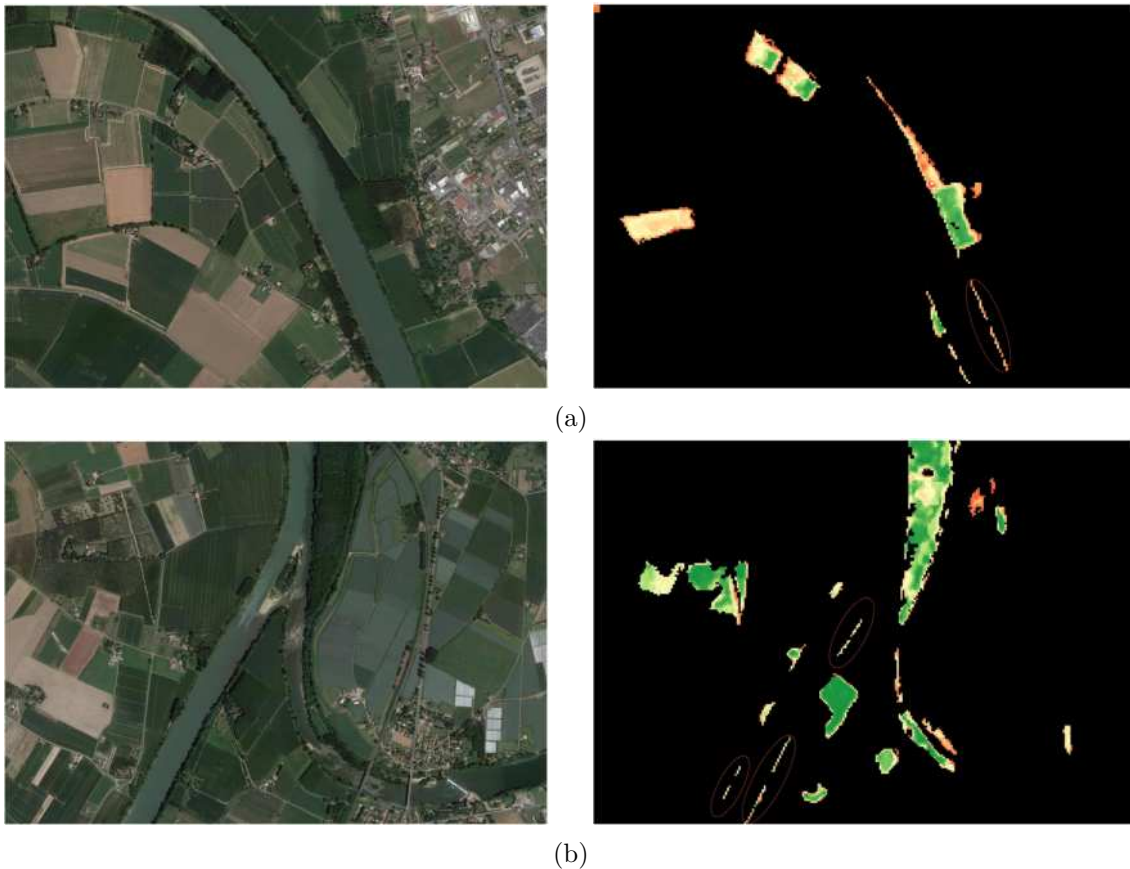


FIGURE 8.2 – Exemples de peupliers naturels détectés en bordure de cours d'eau avec la valeur de confiance associée du rouge (faible confiance) au vert (forte confiance). À gauche, les fonds Google Earth et à droite les confiances associées aux résultats cartographiques pour l'année 2018. Tous les pixels non détectés en peupliers sont masqués en noir.

Ce résultat montre que, même s'il existe des différences structurelles et physiologiques entre les plantations de peupliers et les peupliers naturels, leur comportement spectral n'est pas pour autant très différent. Cependant, l'évaluation des surfaces plantées est sur-estimée puisqu'elle intègre aussi bien les peupleraies que les peupliers naturels. Un travail spécifique est à prévoir en s'aidant de contraintes contextuelle et morphologique pour les distinguer si l'information spectrale est insuffisante.

8.3.3 Validation sur le terrain

Un protocole de validation sur le terrain est indispensable à mettre en place. Cela permettrait d'une part de confirmer les premières interprétations cartographiques mais aussi d'identifier d'éventuelles sur-détections ou sous-détections principalement dans les régions populières où il n'y a pas eu de références pour l'entraînement.

En outre, cette validation permettrait de vérifier la présence potentielle de confusions avec

d'autres espèces non prises en compte dans l'entraînement, notamment les saules (genre *Salix*) qui appartiennent à la même famille que les peupliers (Salicaceae) et qui sont généralement situés près des berges des rivières et des cours d'eau. Bien que d'autres travaux montrent des profils spectraux différents entre les saules en contexte naturel et les peupleraies (Karasiak et al., 2019), ce comportement doit être évalué avec le PI_2 .

Par ailleurs, ce travail de validation sur le terrain permettrait de comprendre l'effet du sous-étage et de déterminer son impact en termes de densité et de composition sur la capacité de détection, principalement dans les peupleraies où il est bien développé.

8.3.4 Dépendance d'un masque forêt

Nous avons souligné dans les Chapitres 5 et 6 l'importance d'utiliser un masque forêt/non forêt afin de limiter les sur-détections des peupleraies. Ces sur-détections concernaient principalement des parcelles agricoles ainsi que certaines plantations d'eucalyptus voire de conifères. Ces prédictions ne sont pas très surprenantes puisque ces classes n'ont pas été considérées lors de l'apprentissage. En outre, nous avons fait le choix de focaliser l'apprentissage sur la reconnaissance des peupleraies et leur distinction des autres classes de feuillus en supposant l'existence d'un masque forêt à l'échelle nationale.

La première piste était de considérer le masque issu de la BD Forêt. Cependant, nous avons constaté que certaines peupleraies détectées peuvent se trouver en dehors du masque. Par ailleurs, certaines zones dont la vocation forestière a disparu résident encore dans le masque de la BD Forêt. En choisissant cette source, il y a un risque de trop sous-estimer ou sur-estimer les surfaces en peupleraies non pas à cause de la qualité de la classification mais à cause de l'inexactitude du masque.

Nous avons également envisagé un masque dérivé de la carte d'occupation des sols du CES OSO. Cependant, la qualité observée du produit s'est avérée insuffisante par rapport à la précision attendue du produit peuplier quant à l'estimation des surfaces. Nous avons principalement constaté des omissions des peupleraies détectées. De plus, ce masque est basé sur des références provenant de la même source que la BD Forêt, ce qui peut induire un biais supplémentaire. Cette source a été délaissée en raison de son inadéquation aux exigences attendues.

Nous nous sommes également intéressés à la création d'un masque forêt à partir des séries temporelles RADAR issues de Sentinel-1 basé sur la moyenne temporelle des images d'hiver et inspiré des travaux de Dostálová et al. (2016, 2018b) et basé sur la forte rétrodiffusion RADAR des forêts en hiver comparée à celle des autres occupations des sols. Un exemple d'un premier masque généré en 2017 est présenté dans la Figure 8.3 où les forêts se démarquent parfaitement des autres formations végétales. Malgré les premiers résultats très prometteurs,

nous ne pouvons pas conclure à ce stade quant à l'efficacité du masque. Ce point mérite d'être davantage investigué.

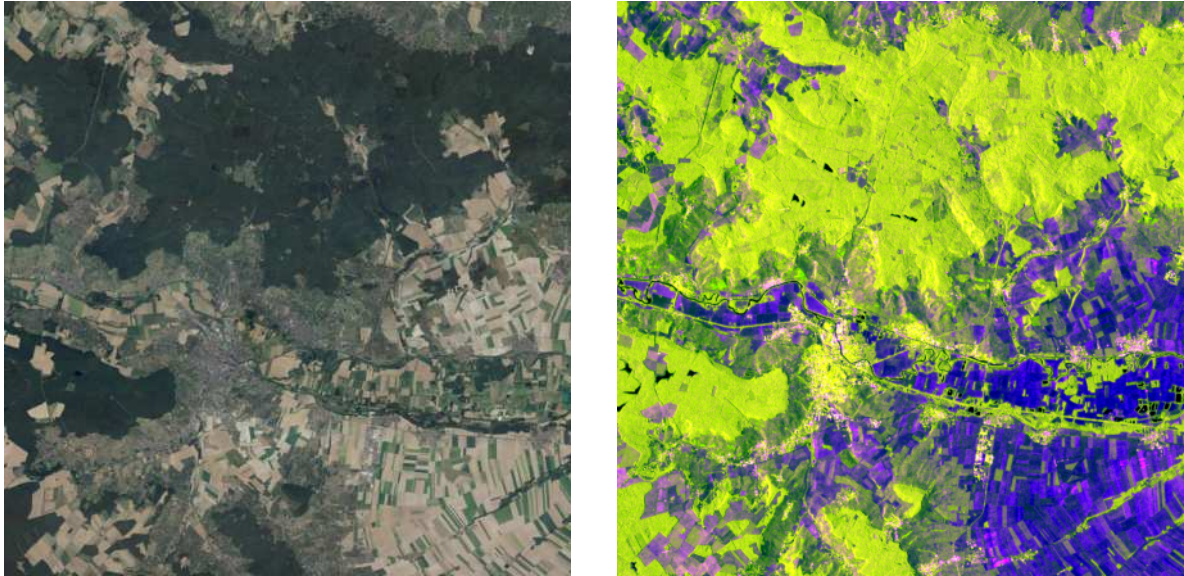


FIGURE 8.3 – Illustration du masque forêt généré à partir de la moyenne temporelle hivernale des images Sentinel-1 de 2017 sur la tuile 31TCJ. À gauche, la photographie aérienne de l'année 2017 et à droite une composition colorées avec en rouge la rétrodiffusion en polarisation VH, en vert la rétrodiffusion en polarisation VV et en bleu le rapport des deux rétrodiffusions.

Enfin, nous avons opté pour le masque HRL de l'année 2018 fourni par l'EEA (Section 7.4.3, point 7). Nous avons jugé ce produit comme étant le meilleur candidat à l'heure actuelle pour masquer les zones hors forêt avec en plus, une cohérence dans sa date de réalisation avec celle de la carte nationale des peupleraies (année 2018).

Il convient maintenant de mieux évaluer l'impact de la qualité de ce masque sur la carte nationale des peupleraies. Cela passe par l'examen des probabilités d'appartenance calculées par l'algorithme. Un travail de validation sur le terrain est également nécessaire.

Au-delà de l'application d'un masque forêt, d'autres pistes ont émergé pour limiter les sur-détections. Il s'agit d'exploiter les valeurs de confiance associées aux résultats de la classification supervisée et de fixer ainsi un seuil de confiance pour la détection. Cela peut être pertinent pour masquer les terres agricoles (en l'absence de masque forêt), dont les valeurs de confiance sont généralement assez faibles, mais un tel seuil peut être défavorable à l'identification de jeunes plantations, pour lesquelles le classifieur est généralement peu confiant.

Un autre critère pertinent est l'information relative à la pente. Les peupleraies étant généralement plantées dans des vallées à faible pente, ce critère peut être intégré comme une variable explicative dans le modèle voire bien *a posteriori* comme un post-traitement visant à éliminer la détection dans les zones très pentues.

8.3.5 Calcul des surfaces et post-traitements

À l'issue du Chapitre 6, la première version de la carte nationale des peupleraies a été générée avec l'indice PI_2 au format raster pour l'année 2018. Il s'agit de la mosaïque des classifications individuelles des 90 tuiles couvrant l'ensemble de la France, masquée par le produit Copernicus HRL forêt de l'EEA. Pour chaque commune, nous avons calculé la densité des plantations en divisant leur surface par celle de la commune, la surface des peupleraies étant calculée par simple comptage de pixels. Nous avons calculé également une valeur médiane de la confiance afin d'évaluer le niveau de certitude du modèle vis-à-vis des prédictions. Les deux variables ont été combinées au sein de la carte choroplèthe de la Figure 8.4.

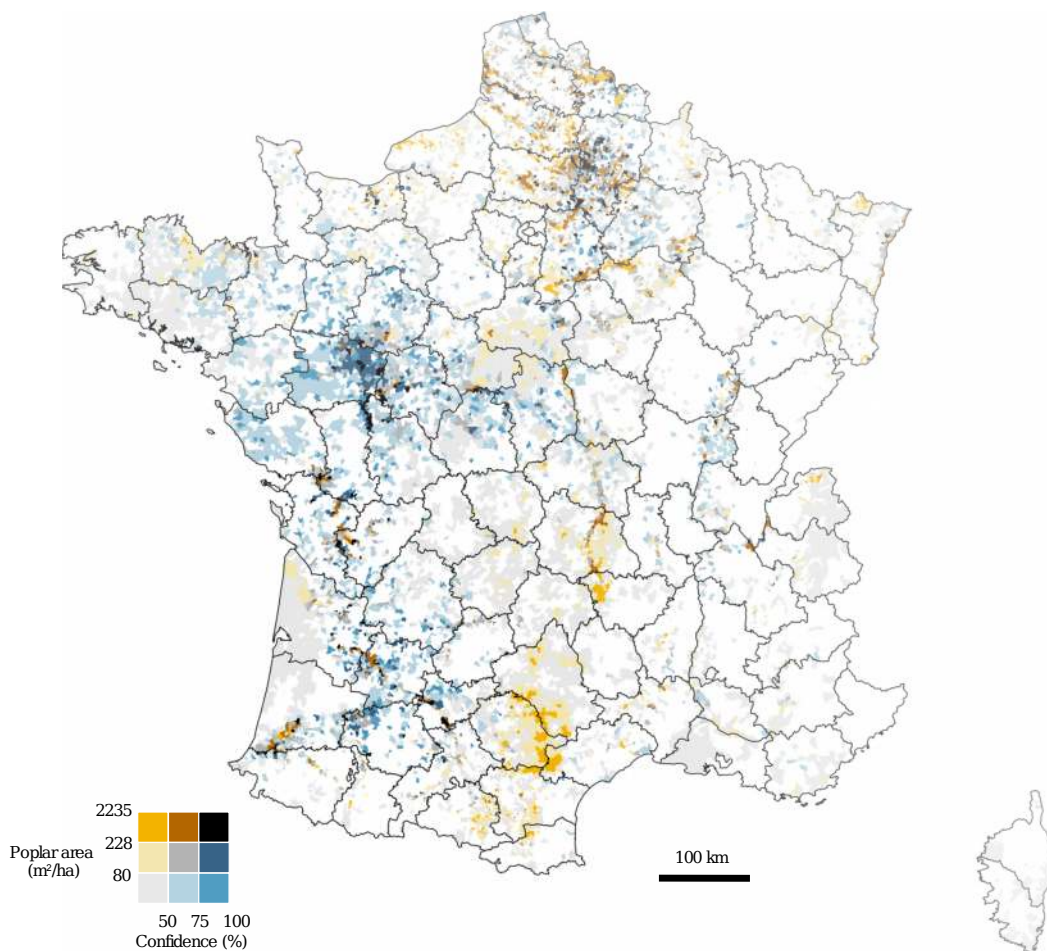


FIGURE 8.4 – Carte choroplèthe bivariée de la France présentant la densité des peupleraies dans chaque commune avec les valeurs médianes de la confiance des prédictions pour l'année 2018.

Cette carte a permis tout d'abord de mettre en évidence les trois bassins populicoles des régions Grand-Est, Centre et Sud-Ouest avec des densités moyennes à élevées associées généralement à de fortes valeurs de confiance. D'autre part, des communes avec une densité élevée et de faibles valeurs de confiances ont également été observées. Il s'agit notamment des trois communes figurant en jaune sur le contour Est du département du Tarn de la région

Sud-Ouest. Les premières interprétations révèlent une sur-détection des peupleraies due aux effets du relief avec des valeurs de pente supérieures à 15%, un cas qui peut être traité en incluant l'information de la pente évoquée dans la section précédente.

Par ailleurs, des calculs de surfaces à l'échelle des parcelles cadastrales est envisagé. Cela impliquera l'application de traitements de généralisation de l'information pixelisée à l'échelle de la parcelle permettant ainsi un suivi à la parcelle mais aussi une estimation plus fine des surfaces. En outre, cette généralisation au niveau des parcelles cadastrales permettra d'obtenir une version vectorielle de la carte des peupleraies. Les premières observations montrent que les limites cadastrales sont parfois trop grossières (polygones rouges en Figure 8.5) et que selon le calcul considéré (nombre de pixels de peupliers, valeur majoritaire, etc.) les estimations de surfaces seront impactées. Les parcelles cadastrales pourraient donc servir à améliorer l'estimation des surfaces mais en limitant leur usage aux cas des peupleraies qui y sont présentes majoritairement.



FIGURE 8.5 – Superposition des limites des parcelles cadastrales (lignes noires) sur les prédictions en peupleraies (pixels verts). Les deux parcelles délimitées en rouge montrent deux exemples de limites cadastrales imprécises.

Pour la production cartographique, nous avons vu que, bien que l'objectif de produire une carte nationale des peupleraies ait été atteint, il reste encore des post-traitements à effectuer, des biais à évaluer quant à la qualité du masque ainsi qu'une validation sur le terrain qui détermineront plus précisément la valeur finale des surfaces.

8.4 Stabilité de la production nationale

Pour cette première production, la chaîne *iota*² a été configurée pour une classification supervisée à partir du PI_2 calculé sur les séries temporelles Sentinel-2 de l'année 2018 et avec des données de référence incluant des polygones de peupleraies mis à jour par photo-interprétation pour la même année. Afin de poursuivre la production sur les prochaines années,

le principe consiste à conserver la même configuration de la chaîne en actualisant à chaque fois les deux entrées : les références des peupleraies et les images Sentinel-2.

Il est envisagé dans un premier temps de lancer la production sur les années 2019 et 2020. Le but étant de confirmer la carte de l'année 2018 et d'analyser les changements d'une année à l'autre, en identifiant si les nouvelles détections sont relatives à des jeunes plantations qui commencent à être identifiées et si les disparitions sont liées à des coupes. Cela permettra également d'identifier les problèmes survenus sur une année ou qui sont localisés dans une zone particulière en raison d'une couverture nuageuse persistante par exemple.

D'un point de vue opérationnel, un protocole de mise à jour des références sur le terrain sera fourni au CNP. Il s'agit principalement de partir d'une base de références peupleraies d'une année N-1, et de vérifier si les parcelles sont toujours valables pour toute l'année N. La diffusion annuelle de la carte des peupleraies est prévue par la plateforme Theia dans le cadre du CES OSO.

Bien que jusqu'à présent, nous ayons envisagé une production sur une base annuelle avec les acquisitions de l'année en question, il serait également possible d'entraîner un modèle en intégrant des données sur une période de deux à trois ans, afin de s'affranchir de l'effet éventuel d'une seule année et de renforcer la qualité de l'état de référence (première carte avant mises à jour successives).

Cela nous ramène également à un autre aspect qui a été peu abordé dans cette thèse à savoir, l'adaptation de domaine dans un contexte de transfert temporel. Dans le Chapitre 4, nous avons brièvement exploré le transfert temporel dans le cadre de l'adaptation des distributions par transport optimal afin d'envisager un transfert d'une année à une autre. Les résultats sur la tuile du Sud-Ouest ont montré généralement une stabilité des prédictions entre deux années consécutives (2017 et 2018) pour les peupleraies sans même faire appel à une adaptation de domaine. En effet, cela implique que le comportement spectral et temporel des peupleraies soit stable d'une année sur l'autre en considérant les mêmes dates. Si la signature spectrale du PI_2 semble être cohérente entre les deux années testées (Figure B.3), il convient de vérifier ce comportement sur plusieurs années.

Chapitre 9

Conclusion générale

Sommaire

| | | |
|-----|------------------------|------------|
| 9.1 | Conclusions | 206 |
| 9.2 | Perspectives | 207 |

9.1 Conclusions

Les travaux de cette thèse ont porté sur la reconnaissance automatique des peupleraies à partir de données hypertextuelles de télédétection optique, en réponse au besoin de la filière peuplier de disposer d'un outil de suivi régulier de cette ressource aux enjeux économiques et sociaux importants.

Le contexte de disponibilité en images satellitaires à moindre coût était favorable et pertinent pour réinterroger les capacités de la télédétection à relever ce défi. L'arrivée des nouvelles constellations de satellites telles que Sentinel-2 a permis depuis quelques années de fournir des acquisitions avec une couverture globale et à hautes résolutions spatiale, spectrale et temporelle.

Le principal objectif de cette thèse était ainsi d'étudier le potentiel des séries temporelles Sentinel-2 pour l'identification des peupleraies à large échelle. Plus particulièrement, il s'agissait de développer une approche méthodologique opérationnelle permettant la production régulière d'une carte nationale des peupleraies françaises.

Pour y répondre, nous avons entrepris deux principales stratégies, la première basée sur deux techniques d'adaptation de domaine, peu utilisées en télédétection, pour la généralisation spatiale des connaissances tout en limitant le nombre d'échantillons de référence, et la seconde, plus classique, visant à identifier les caractéristiques spectrales et temporelles les plus pertinentes pour discriminer les peupleraies dans les différents contextes populeux nationaux mais qui exploite l'ensemble des références disponibles.

Nous avons montré l'intérêt de chacune des trois approches proposées pour l'identification des peupleraies à large échelle avec toutefois des degrés de complexité différents pour une mise en œuvre opérationnelle.

Dans le cas de l'adaptation de domaine par transport optimal ou par apprentissage actif, nous avons mis en évidence leurs efficacités en termes de réduction du nombre d'échantillons de référence en tirant parti des connaissances acquises précédemment. Nous avons par ailleurs apporté une meilleure compréhension de leurs fonctionnements en considérant à la fois l'information spectrale et temporelle contenues dans les données satellitaires. Cependant la relative complexité des algorithmes d'adaptation de domaine et de leur déploiement à l'échelle nationale rendent leur transfert vers des partenaires non spécialistes assez difficile.

La troisième approche a permis de proposer un nouvel indice spécifique aux peupliers, le PI_2 , qui a été utilisé pour la production de la première carte des peupleraies en France. Il s'agit d'une méthode robuste, simple et spécifique aux peupliers, assurant à la filière un outil opérationnel assurant une production régulière. Si cette approche a été retenue pour le passage à l'échelle nationale, elle reste très exigeante en termes de références qui nécessitent

d’être mises à jour en vue d’une production annuelle régulière. Un enjeu important de cette production cartographique est l’estimation des surfaces de peupleraies, qui constitue l’une des principales attentes de la filière. La précision des estimations repose sur la qualité de la carte, qui dépend elle-même de plusieurs facteurs, dont la qualité du masque forêt et les opérations de simplification géométrique.

Au regard de ces limites, cette thèse ouvre de nouvelles perspectives méthodologique et opérationnelle qui seront abordées dans la section suivante.

9.2 Perspectives

Ce travail de thèse qui porte sur le développement d’une approche opérationnelle pour l’identification des peupleraies à large échelle débouche sur de nombreuses perspectives.

À court terme, il s’agit d’abord de lancer la campagne de validation de la première carte nationale de 2018 sur le terrain. Pour cela un plan d’échantillonnage doit être mis en place en focalisant la validation principalement sur les zones non prises comme référence pour l’entraînement du modèle. Ces retours permettront de mieux qualifier et quantifier les erreurs afin de pouvoir y remédier.

Dans un second temps, il convient de produire d’autres millésimes de la carte des peupleraies avec la méthode basée sur le PI_2 retenue pour le passage à l’échelle. Cela permettra une comparaison entre années afin d’identifier les jeunes plantations qui devraient devenir de plus en plus détectables avec le développement de leurs canopées, mais aussi les disparitions dues aux coupes d’une année à l’autre. Des méthodes de détection des changements annuels pourront ainsi être mises en place.

D’un point de vue méthodologique, les approches d’adaptation de domaine gagneraient à une réduction de la dimension. Il serait pertinent de tirer parti des résultats de la troisième approche en évaluant la contribution des deux techniques d’adaptation de domaine sur les séries temporelles du PI_2 au lieu de considérer toute la dimension spectrale des données Sentinel-2. Par ailleurs, les trois approches proposées peuvent être combinées. Il serait ainsi possible d’appliquer le transport optimal en mode non supervisé sur les séries temporelles du PI_2 et de l’améliorer en rajoutant des échantillons par apprentissage actif afin d’aboutir à une carte nationale.

D’autre part, il serait envisageable de construire des modèles pluriannuels au lieu de l’approche annuelle proposée. Cela rendrait le modèle plus robuste, moins dépendant de l’effet d’une année particulière, et éviterait également de devoir fournir chaque année des références actualisées sur les peupleraies. Par ailleurs, nous avons vu que le comportement temporel du PI_2 était assez similaire entre les deux années d’étude. Cet aspect doit être davantage approfondi et vérifié entre plusieurs années afin de confirmer si le PI_2 est finalement stable dans le

temps pour les peupleraies. Il est ainsi possible qu'un modèle entraîné sur une année donnée soit également valable pour prédire les années suivantes. Cependant, comme le PI_2 met principalement en évidence la teneur en eau des feuilles, la réponse spectrale peut être différente lors d'une année particulièrement plus sèche que celle des données utilisées pour l'entraînement, ce qui constitue tout de même la tendance actuelle dans un contexte de dérèglement climatique. Il est également possible que même si la réponse spectrale n'est pas identique d'une année à l'autre et qu'elle est influencée par les conditions climatiques, la variation demeurera relative et l'écart entre les peupliers et les autres espèces sera maintenu.

D'un point de vue cartographique, nous avons vu que indépendamment du masque forêt choisi, il y a toujours un biais associé et les résultats finaux dépendent de sa qualité, de même que l'estimation des surfaces. Deux approches sont ainsi envisageables. La première consiste à créer un masque forêt adapté et à approfondir les premières expérimentations encourageantes avec l'utilisation de la moyenne saisonnière des données Sentinel-1 pour séparer les forêts des autres occupations des sols. Ceci est d'autant plus pertinent que les données Sentinel-1 sont désormais supportées par la chaîne de traitement *iota*². La seconde approche consiste à rendre la production complètement indépendante d'un masque forêt, en considérant d'autres classes d'occupation des sols dans l'apprentissage comme les conifères ou les parcelles agricoles. Une approche de classification avec différents niveaux thématiques allant des grandes catégories (forêt, urbain, eau, etc.) jusqu'au niveau de l'essence peut ainsi être envisagée.

D'autre part, il est également envisagé d'utiliser les limites des parcelles cadastrales afin de généraliser les résultats cartographiques à la parcelle et d'obtenir une version vectorielle des cartes au format raster. Cela permettrait d'avoir un rendu plus homogène (non pixelisé) et d'affiner ainsi l'estimation des surfaces, qui demeure la principale préoccupation des partenaires de la filière. En plus des résultats de la classification, ces généralisations cartographiques peuvent prendre en compte les valeurs de confiance pour établir un intervalle d'erreur pour les estimations. Compte tenu du besoin spécifique du produit peuplier, il paraît peu envisageable d'intégrer cette étape directement dans *iota*² mais plutôt de la concevoir sous la forme d'une boîte à outils dans un système d'information géographique à destination des utilisateurs potentiels.

Références bibliographiques

Références bibliographiques

- Alajlan, N., Pasoli, E., Melgani, F., and Franzoso, A. (2014). Large-Scale Image Classification Using Active Learning. *IEEE Geoscience and Remote Sensing Letters*, 11(1) :259–263. [34](#), [116](#), [137](#)
- Almeida, A. E. (2017). Time series components and breakpoints in remote sensing image analysis = Componentes e pontos de quebra em séries temporais na análise de imagens de sensoriamento remoto. *undefined*. [38](#)
- Amor, I. B. S. B., Chehata, N., Bailly, J., Farah, I. R., and Lagacherie, P. (2018). Parcel-Based Active Learning for Large Extent Cultivated Area Mapping. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(1) :79–88. [116](#)
- Angluin, D. (1988). Queries and Concept Learning. *Machine Learning*, 2(4) :319–342. [124](#)
- Apan, A., Held, A., Phinn, S., and Markley, J. (2003). Formulation and assessment of narrow-band vegetation indices from EO-1 hyperion imagery for discriminating sugarcane disease. In *Proceedings of the Spatial Sciences Institute Biennial Conference (SSC 2003) : Spatial Knowledge Without Boundaries*, pages 1–13, Canberra, Australia. Spatial Sciences Institute. [159](#)
- Asner, G. P. and Alencar, A. (2010). Drought impacts on the Amazon forest : The remote sensing perspective : Research review. *New Phytologist*, 187(3) :569–578. [22](#)
- Atkinson, P. M., Jeganathan, C., Dash, J., and Atzberger, C. (2012). Inter-comparison of four models for smoothing satellite sensor time-series data to estimate vegetation phenology. *Remote Sensing of Environment*, 123 :400–417. [49](#)
- Atzberger, C. (2013). Advances in Remote Sensing of Agriculture : Context Description, Existing Operational Monitoring Systems and Major Information Needs. *Remote Sensing*, 5(2) :949–981. [28](#)
- Axelsson, A., Lindberg, E., Reese, H., and Olsson, H. (2021). Tree species classification using Sentinel-2 imagery and Bayesian inference. *International Journal of Applied Earth Observation and Geoinformation*, 100 :102318. [24](#)
- Baetens, L., Desjardins, C., and Hagolle, O. (2019). Validation of Copernicus Sentinel-2 Cloud Masks Obtained from MAJA, Sen2Cor, and FMask Processors Using Reference Cloud Masks Generated with a Supervised Active Learning Procedure. *Remote Sensing*, 11(4) :433. [45](#)
- Balzter, H., Cole, B., Thiel, C., and Schmullius, C. (2015). Mapping CORINE Land Cover from Sentinel-1A SAR and SRTM Digital Elevation Model Data using Random Forests. *Remote Sensing*, 7(11) :14876–14898. [61](#)

-
- Banskota, A., Kayastha, N., Falkowski, M. J., Wulder, M. A., Froese, R. E., and White, J. C. (2014). Forest Monitoring Using Landsat Time Series Data : A Review. *Canadian Journal of Remote Sensing*, 40(5) :362–384. [23](#)
- Belgiu, M. and Drăguț, L. (2016). Random forest in remote sensing : A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114 :24–31. [64](#), [123](#)
- Bergante, S., Facciotto, G., and Minotta, G. (2010). Identification of the main site factors and management intensity affecting the establishment of Short-Rotation-Coppices (SRC) in Northern Italy through stepwise regression analysis. *Open Life Sciences*, 5(4) :522–530. [173](#)
- Bolyn, C., Latte, N., Colson, V., Fourbisseur, A., Vanderheeren, N., and Lejeune, P. (2020). Une carte des principaux types de peuplements forestiers de Belgique et du Nord de la France. *Forêt.Nature*, (156) :48 – 57. [26](#)
- Bolyn, C., Michez, A., Gaucher, P., Lejeune, P., and Bonnet, S. (2018). Forest mapping and species composition using supervised per pixel classification of Sentinel-2 imagery. *Biotechnology, Agronomy, Society and Environment*. [24](#)
- Borry, F. C., de Roover, B. P., Leysen, M. M., de Wulf, R. R., and Goossens, R. E. (1993). Evaluation of SPOT and TM data for forest stratification : A case study for small-size poplar stands. *IEEE Transactions on Geoscience and Remote Sensing*, 31(2) :483–490. [26](#), [113](#), [150](#)
- Botkin, D. B., Estes, J. E., MacDonald, R. M., and Wilson, M. V. (1984). Studying the Earth’s Vegetation from Space. *BioScience*, 34(8) :508–514. [32](#)
- Boyd, D.S.Danson, F. (2005). Satellite remote sensing of forest resources : three decades of research development. *Progress in Physical Geography*, 29(1) :1–26. [13](#), [149](#)
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1) :5–32. [61](#), [62](#), [122](#), [123](#), [156](#)
- Brinker, K. (2003). Incorporating Diversity in Active Learning with Support Vector Machines. In *International Conference on Machine Learning*, pages 59–66, Washington, DC, USA. AAAI Press. [125](#)
- Bruzzone, L. and Carlin, L. (2006). A Multilevel Context-Based System for Classification of Very High Spatial Resolution Images. *IEEE Transactions on Geoscience and Remote Sensing*, 44(9) :2587–2600. [29](#)
- Bruzzone, L., Chi, M., and Marconcini, M. (2006). A Novel Transductive SVM for Semisupervised Classification of Remote-Sensing Images. *IEEE Transactions on Geoscience and Remote Sensing*, 44(11) :3363–3373. [33](#)
- Bruzzone, L. and Cossu, R. (2002). A multiple-cascade-classifier system for a robust and partially unsupervised updating of land-cover maps. *IEEE Transactions on Geoscience and Remote Sensing*, 40(9) :1984–1996. [116](#)
- Bruzzone, L. and Marconcini, M. (2009). Toward the Automatic Updating of Land-Cover Maps by a Domain-Adaptation SVM Classifier and a Circular Validation Strategy. *IEEE Transactions on Geoscience and Remote Sensing*, 47(4) :1108–1122. [34](#)
- Bruzzone, L. and Persello, C. (2009). A Novel Approach to the Selection of Spatially Invariant Features for the Classification of Hyperspectral Images With Improved Generalization Capability. *IEEE Transactions on Geoscience and Remote Sensing*, 47(9) :3180–3191. [34](#)



-
- Bruzzone, L. and Prieto, D. (2001). Unsupervised retraining of a maximum likelihood classifier for the analysis of multitemporal remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 39(2) :456–460. [116](#)
- Cano, E., Denux, J.-P., Bisquert, M., Hubert-Moy, L., and Chéret, V. (2017). Improved forest-cover mapping based on MODIS time series and landscape stratification. *International Journal of Remote Sensing*, 38(7) :1865–1888. [23](#)
- Cao, X., Yao, J., Xu, Z., and Meng, D. (2020). Hyperspectral image classification with convolutional neural network and active learning. *IEEE Transactions on Geoscience and Remote Sensing*, 58(7) :4604–4616. [116](#)
- Carah Asbl (2019). Brochure peuplier & populiculture 2.0. <https://www.foret-probos.eu/fr/publications/publication/69:brochure-peuplier-populiculture-2-0>. [7](#)
- Carle, J. and Holmgren, P. (2008). Wood from planted forests. *Forest Products Journal*, 58(12) :6. [149](#)
- Cassisi, C., Montalto, P., Aliotta, M., Cannata, A., and Pulvirenti, A. (2012). *Similarity Measures and Dimensionality Reduction Techniques for Time Series Data Mining*. IntechOpen. [108](#)
- Chan, J. C.-W., Beckers, P., Spanhove, T., and Borre, J. V. (2012). An evaluation of ensemble classifiers for mapping Natura 2000 heathland in Belgium using spaceborne angular hyperspectral (CHRIS/Proba) imagery. *International Journal of Applied Earth Observation and Geoinformation*, 18 :13–22. [61](#)
- Chandrashekar, G. and Sahin, F. (2014). A survey on feature selection methods. *Computers & Electrical Engineering*, 40(1) :16–28. [154](#), [155](#)
- Chardenon, J. and Flouzat, G. (1981). The application of remote sensing to poplar growing : identification and inventory of poplar groves, prediction of timber production ; France, Italy. *Revue Forestiere Francaise*, 33(6) :478–493. [25](#), [26](#), [113](#), [150](#)
- Chen, Y. and Gong, P. (2013). Clustering based on eigenspace transformation – CBEST for efficient classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 83 :64–80. [30](#), [57](#)
- Cheng, Y., Yu, L., Xu, Y., Liu, X., Lu, H., Cracknell, A. P., Kanniah, K., and Gong, P. (2018). Towards global oil palm plantation mapping using remote-sensing data. *International Journal of Remote Sensing*, 39(18) :5891–5906. [25](#), [113](#)
- Cibula, W. G., Zetka, E. F., and Rickman, D. L. (1992). Response of thematic mapper bands to plant water stress. *International Journal of Remote Sensing*, 13(10) :1869–1880. [245](#)
- Clark, M. L. and Roberts, D. A. (2012). Species-Level Differences in Hyperspectral Metrics among Tropical Rainforest Trees as Determined by a Tree-Based Classifier. *Remote Sensing*, 4(6) :1820–1855. [19](#)
- Clevers, J. G. P. W., Kooistra, L., and Schaepman, M. E. (2008). Using spectral information from the NIR water absorption features for the retrieval of canopy water content. *International Journal of Applied Earth Observation and Geoinformation*, 10(3) :388–397. [171](#)
- Cloutis, E. A., Connery, D. R., Major, D. J., and Dover, F. J. (1996). Airborne multi-spectral monitoring of agricultural crop status : Effect of time of year, crop type and crop condition parameter. *International Journal of Remote Sensing*, 17(13) :2579–2601. [159](#)

-
- CNES (2021). Formation nouveaux utilisateurs – HAL. [181](#)
- Cohen, W. B., Maiersperger, T. K., Gower, S. T., and Turner, D. P. (2003). An improved strategy for regression of biophysical variables and Landsat ETM+ data. *Remote Sensing of Environment*, 84(4) :561–571. [23](#)
- Cohn, D., Atlas, L., and Ladner, R. (1994). Improving generalization with active learning. *Machine Learning*, 15(2) :201–221. [124](#)
- Conseil National du Peuplier (2019). Sécurisation des approvisionnements en Peuplier : Étude prospective de la ressource française. Technical report, Union des Industries du Panneau Contreplaqué et Comité professionnel de Développement des Industries Françaises de l’Ameublement et du Bois. [8](#), [9](#)
- Courty, N., Flamary, R., and Tuia, D. (2014). Domain Adaptation with Regularized Optimal Transport. In Calders, T., Esposito, F., Hüllermeier, E., and Meo, R., editors, *Machine Learning and Knowledge Discovery in Databases*, Lecture Notes in Computer Science, pages 274–289, Berlin, Heidelberg. Springer. [74](#), [75](#), [76](#), [79](#), [82](#), [108](#)
- Crawford, M. M., Tuia, D., and Yang, H. L. (2013). Active Learning : Any Value for Classification of Remotely Sensed Data? *Proceedings of the IEEE*, 101(3) :593–608. [114](#), [115](#), [140](#)
- Curran, P. J. (1989). Remote sensing of foliar chemistry. *Remote Sensing of Environment*, 30(3) :271–278. [171](#)
- Cutler, D. R., Edwards, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., and Lawler, J. J. (2007). Random Forests for Classification in Ecology. *Ecology*, 88(11) :2783–2792. [61](#)
- Cuturi, M. (2013). Sinkhorn Distances : Lightspeed Computation of Optimal Transport. *Advances in Neural Information Processing Systems*, 26. [76](#)
- Cuturi, M. and Blondel, M. (2017). Soft-DTW : A Differentiable Loss Function for Time-Series. In *International Conference on Machine Learning*, pages 894–903. PMLR. [108](#)
- Dalponte, M., Ørka, H. O., Gobakken, T., Gianelle, D., and Næsset, E. (2013). Tree Species Classification in Boreal Forests With Hyperspectral Data. *IEEE Transactions on Geoscience and Remote Sensing*, 51(5) :2632–2645. [61](#)
- Danka, T. and Horvath, P. (2018). modAL : A modular active learning framework for Python. [126](#)
- Dannenberg, M. P., Hakkenberg, C. R., and Song, C. (2016). Consistent Classification of Landsat Time Series with an Improved Automatic Adaptive Signature Generalization Algorithm. *Remote Sensing*, 8(8) :691. [32](#)
- Dash, J. and Curran, P. J. (2004). The MERIS terrestrial chlorophyll index. *International Journal of Remote Sensing*, 25(23) :5403–5413. [171](#)
- Dask Development Team (2016). *Dask : Library for dynamic task scheduling*. [179](#)
- Datt, B. (1999). Remote Sensing of Water Content in Eucalyptus Leaves. *Australian Journal of Botany*, 47(6) :909–923. [159](#)
- Demir, B., Minello, L., and Bruzzone, L. (2014). Definition of Effective Training Sets for Supervised Classification of Remote Sensing Images by a Novel Cost-Sensitive Active Learning Method. *IEEE Transactions on Geoscience and Remote Sensing*, 52(2) :1272–1284. [116](#)

-
- Demir, B., Persello, C., and Bruzzone, L. (2011). Batch-mode active-learning methods for the interactive classification of remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 49(3) :1014–1031. [116](#)
- Deng, X., Guo, S., Sun, L., and Chen, J. (2020). Identification of Short-Rotation Eucalyptus Plantation at Large Scale Using Multi-Satellite Imageries and Cloud Computing Platform. *Remote Sensing*, 12(13) :2153. [25](#), [149](#)
- Descals, A., Szantoi, Z., Meijaard, E., Sutikno, H., Rindanata, G., and Wich, S. (2019). Oil Palm (*Elaeis guineensis*) Mapping with Details : Smallholder versus Industrial Plantations and their Extent in Riau, Sumatra. *Remote Sensing*, 11(21) :2590. [25](#), [113](#), [149](#)
- Devineau, J.-L. (1990). Propriétés spectrales de la végétation. *Images satellite et milieux terrestres en régions arides et tropicales*, pages 43–49. [18](#)
- Di, W. and Crawford, M. M. (2011). Critical class oriented active learning for hyperspectral image classification. In *2011 IEEE International Geoscience and Remote Sensing Symposium*, pages 3899–3902, Vancouver, BC, Canada. IEEE. [138](#), [140](#)
- Dickmann, D. I. (2006). Silviculture and biology of short-rotation woody crops in temperate regions : Then and now. *Biomass and Bioenergy*, 30(8) :696–705. [4](#), [6](#)
- Dietterich, T. G. (1998). Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Computation*, 10(7) :1895–1923. [156](#)
- Dong, J., Xiao, X., Sheldon, S., Biradar, C., and Xie, G. (2012). Mapping tropical forests and rubber plantations in complex landscapes by integrating PALSAR and MODIS imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 74 :20–33. [23](#), [113](#)
- Dong, T., Liu, J., Shang, J., Qian, B., Ma, B., Kovacs, J. M., Walters, D., Jiao, X., Geng, X., and Shi, Y. (2019). Assessment of red-edge vegetation indices for crop leaf area index estimation. *Remote Sensing of Environment*, 222 :133–143. [171](#)
- Dostálová, A., Hollaus, M., Milenković, M., and Wagner, W. (2016). FOREST AREA DERIVATION FROM SENTINEL-1 DATA. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, III-7 :227–233. [200](#)
- Dostálová, A., Wagner, W., Milenković, M., and Hollaus, M. (2018a). Annual seasonality in Sentinel-1 signal for forest mapping and forest type classification. *International Journal of Remote Sensing*, 39(21) :7738–7760. [24](#)
- Dostálová, A., Wagner, W., Milenković, M., and Hollaus, M. (2018b). Annual seasonality in Sentinel-1 signal for forest mapping and forest type classification. *International Journal of Remote Sensing*, 0(0) :1–23. [200](#)
- Dostálová, A., Lang, M., Ivanovs, J., Waser, L. T., and Wagner, W. (2021). European wide forest classification based on sentinel-1 data. *Remote Sensing*, 13(3). [175](#)
- Du Cros, E. T. (1984). Breeding strategies with poplars in Europe. *Forest Ecology and Management*, 8(1) :23–39. [3](#)
- Duchauffour, P. (1955). Les sols à peupliers. *Revue Forestière Française*, (7) :539. [139](#)
- Eckenwalder, J. E. (1996). Taxonomic Signal and Noise in Multivariate Interpopulational Relationships in *Populus mexicana* (Salicaceae). *Systematic Botany*, 21(3) :261–271. [3](#)
- ESA (2015). Sentinel-2 User Handbook. Technical report. [44](#)



-
- Eslami, A. and Zahedi, S. S. (2011). Providing poplar plantation map by Indian remote sensing (IRS) satellite imagery in Northern Iran. *African Journal of Agricultural Research*, Vol. 6(20) :4769–4774. [25](#), [26](#), [113](#), [150](#)
- FAO (2016a). Poplars and Other Fast-Growing Trees - Renewable Resources for Future Green Economies. In *Synthesis of Country Progress Reports*, page 120, Federal Republic of Germany, Berlin. Forestry Policy and Resources Division. [6](#), [7](#), [113](#), [149](#)
- FAO (2016b). The reform of the International Poplar Commission (IPC) in the light of modern forest policy requirements. In *Abstracts of Submitted Papers Prepared for the 25 Th Session of the International Poplar Commission*, page 10, Berlin, Germany. International Poplar Commission Working Paper IPC/14. Forestry Policy and Resources Division. [4](#)
- FAO (2020). *Global Forest Resources Assessment 2020 : Key Findings*. FAO, Rome, Italy. [149](#)
- Fares, S., Mugnozsa, G., Corona, P., and Palahí, M. (2015). Sustainability : Five steps for managing europe’s forests. *Nature*, 519(7544) :407–409. [149](#)
- Fassnacht, F. E., Latifi, H., Stereńczak, K., Modzelewska, A., Lefsky, M., Waser, L. T., Straub, C., and Ghosh, A. (2016). Review of studies on tree species classification from remotely sensed data. *Remote Sensing of Environment*, 186 :64–87. [17](#), [21](#), [24](#), [149](#)
- Fensholt, R. and Sandholt, I. (2003). Derivation of a shortwave infrared water stress index from MODIS near- and shortwave infrared data in a semiarid environment. *Remote Sensing of Environment*, 87(1) :111–121. [159](#)
- Filipponi, F. (2018). BAIS2 : Burned Area Index for Sentinel-2. *Proceedings*, 2(7) :364. [245](#)
- Fischer, M., Zenone, T., Trnka, M., Orság, M., Montagnani, L., Ward, E. J., Tripathi, A. M., Hlavinka, P., Seufert, G., Žalud, Z., King, J. S., and Ceulemans, R. (2018). Water requirements of short rotation poplar coppice : Experimental and modelling analyses across Europe. *Agricultural and Forest Meteorology*, 250-251 :343–360. [139](#)
- Flamary, R., Courty, N., Gramfort, A., Alaya, M. Z., Boisbunon, A., Chambon, S., Chapel, L., Corenflos, A., Fatras, K., Fournier, N., Gautheron, L., Gayraud, N. T., Janati, H., Rakotomamonjy, A., Redko, I., Rolet, A., Schutz, A., Seguy, V., Sutherland, D. J., Tavenard, R., Tong, A., and Vayer, T. (2021). Pot : Python optimal transport. *Journal of Machine Learning Research*, 22(78) :1–8. [85](#)
- Folleco, A., Khoshgoftaar, T. M., Van Hulse, J., and Bullard, L. (2008). Identifying learners robust to low quality data. In *2008 IEEE International Conference on Information Reuse and Integration*, pages 190–195. [196](#)
- Foody, G. M. (2002). Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, 80(1) :185–201. [66](#)
- Foody, G. M. (2004). Thematic map comparison : Evaluating the statistical significance of differences in classification accuracy. *Photogrammetric Engineering and Remote Sensing*, 70(5) :627–633. [156](#)
- Foody, G. M. (2020). Explaining the unsuitability of the kappa coefficient in the assessment and comparison of the accuracy of thematic maps obtained by image classification. *Remote Sensing of Environment*, 239 :111630. [66](#), [67](#), [123](#), [156](#)

-
- Foody, G. M. and Mathur, A. (2006). The use of small training sets containing mixed pixels for accurate hard image classification : Training on mixed spectral responses for classification by a SVM. *Remote Sensing of Environment*, 103(2) :179–189. [31](#)
- Forstmaier, A., Shekhar, A., and Chen, J. (2020). Mapping of Eucalyptus in Natura 2000 Areas Using Sentinel 2 Imagery and Artificial Neural Networks. *Remote Sensing*, 12(14) :2176–25, [149](#)
- Frampton, W. J., Dash, J., Watmough, G., and Milton, E. J. (2013). Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variables in vegetation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 82 :83–92. [245](#)
- Franklin, S. E., Hall, R. J., Moskal, L. M., Maudie, A. J., and Lavigne, M. B. (2000). Incorporating texture into classification of forest species composition from airborne multispectral images. *International Journal of Remote Sensing*, 21(1) :61–79. [24](#)
- Franklin, S. E. and Wulder, M. A. (2002). Remote sensing methods in medium spatial resolution satellite data land cover classification of large areas. *Progress in Physical Geography : Earth and Environment*, 26(2) :173–205. [30](#), [31](#)
- Friedl, M. A., McIver, D. K., Hodges, J. C. F., Zhang, X. Y., Muchoney, D., Strahler, A. H., Woodcock, C. E., Gopal, S., Schneider, A., Cooper, A., Baccini, A., Gao, F., and Schaaf, C. (2002). Global land cover mapping from MODIS : Algorithms and early results. *Remote Sensing of Environment*, 83(1) :287–302. [23](#)
- Frison, P.-L., Fruneau, B., Kmiha, S., Soudani, K., Dufrière, E., Le Toan, T., Koleček, T., Villard, L., Mougin, E., and Rudant, J.-P. (2018). Potential of Sentinel-1 Data for Monitoring Temperate Mixed Forest Phenology. *Remote Sensing*, 10(12) :2049. [24](#)
- Fuller, D. O. (2006). Tropical forest monitoring and remote sensing : A new era of transparency in forest governance? *Singapore Journal of Tropical Geography*, 27(1) :15–29. [19](#)
- Gao, B.-C. (1996). NdwI—a normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment*, 58(3) :257–266. [171](#), [245](#)
- Gao, F., Masek, J., Schwaller, M., and Hall, F. (2006). On the blending of the Landsat and MODIS surface reflectance : Predicting daily Landsat surface reflectance. *IEEE Transactions on Geoscience and Remote Sensing*, 44(8) :2207–2218. [24](#)
- Gao, T., Zhu, J., Zheng, X., Shang, G., Huang, L., and Wu, S. (2015). Mapping Spatial Distribution of Larch Plantations from Multi-Seasonal Landsat-8 OLI Imagery and Multi-Scale Textures Using Random Forests. *Remote Sensing*, 7(2) :1702–1720. [25](#)
- Garbulsky, M. F., Peñuelas, J., Gamon, J., Inoue, Y., and Filella, I. (2011). The photochemical reflectance index (PRI) and the remote sensing of leaf, canopy and ecosystem radiation use efficiencies : A review and meta-analysis. *Remote Sensing of Environment*, 115(2) :281–297. [171](#)
- Ghosh, A., Fassnacht, F. E., Joshi, P. K., and Koch, B. (2014). A framework for mapping tree species combining hyperspectral and LiDAR data : Role of selected classifiers and sensor across three spatial scales. *International Journal of Applied Earth Observation and Geoinformation*, 26 :49–63. [24](#), [61](#)
- Ghosh, A. and Joshi, P. K. (2014). A comparison of selected classification algorithms for mapping bamboo patches in lower Gangetic plains using very high resolution WorldView 2

-
- imagery. *International Journal of Applied Earth Observation and Geoinformation*, 26 :298–311. [64](#), [123](#), [156](#)
- Girard, M. C., Girard, C. M., and Courault, D. (1999). *Traitement des données de télédétection*. Dunod, Paris. OCLC : 300177600. [18](#), [19](#)
- Gitelson, A. A., Merzlyak, M., Zur, Y., Stark, R., and Gritz, U. (2001). Non-destructive and remote sensing techniques for estimation of vegetation status. In *Proceedings of the 3rd European Conference on Precision Agriculture*, pages 205–210. [159](#)
- Gitelson, A. A., Merzlyak, M. N., and Lichtenthaler, H. K. (1996). Detection of Red Edge Position and Chlorophyll Content by Reflectance Measurements Near 700 nm. *Journal of Plant Physiology*, 148(3) :501–508. [171](#), [172](#)
- Gómez, C., White, J. C., and Wulder, M. A. (2016). Optical remotely sensed time series data for land cover classification : A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 116 :55–72. [28](#), [29](#), [30](#), [58](#)
- Gomez-Chova, L., Calpe, J., Camps-Valls, G., Martin, J., Soria, E., Vila, J., Alonso-Chorda, L., and Moreno, J. (2003). Feature selection of hyperspectral data through local correlation and sffs for crop classification. In *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, volume 1, pages 555–557. [155](#)
- Gong, B., Grauman, K., and Sha, F. (2013). Connecting the Dots with Landmarks : Discriminatively Learning Domain-Invariant Features for Unsupervised Domain Adaptation. In *International Conference on Machine Learning*, pages 222–230. [33](#), [115](#)
- Gong, Z., Zhong, P., and Hu, W. (2019). Diversity in Machine Learning. *IEEE Access*, 7 :64323–64350. [125](#)
- Gordon, J. C. (2011). Poplars : Trees of the people, trees of the future. *The Forestry Chronicle*. [3](#)
- Grekousis, G., Mountrakis, G., and Kavouras, M. (2015). An overview of 21 global and 43 regional land-cover mapping products. *International Journal of Remote Sensing*, 36(21) :5309–5335. [30](#)
- Grignetti, A., Coaloa, D., and Niccolini, G. (2009). Classification of poplar stand areas by high-resolution satellite images. *Forest@ - Rivista di Selvicoltura ed Ecologia Forestale*, 6(1) :299–311. [26](#), [113](#), [150](#)
- Guerschman, J. P., Hill, M. J., Renzullo, L. J., Barrett, D. J., Marks, A. S., and Botha, E. J. (2009). Estimating fractional cover of photosynthetic vegetation, non-photosynthetic vegetation and bare soil in the Australian tropical savanna region upscaling the EO-1 Hyperion and MODIS sensors. *Remote Sensing of Environment*, 113(5) :928–945. [159](#)
- Hagolle, O., Huc, M., Desjardins, C., Auer, S., and Richter, R. (2017). MAJA Algorithm Theoretical Basis Document. [45](#), [46](#)
- Hagolle, O., Huc, M., Villa Pascual, D., and Dedieu, G. (2015). A Multi-Temporal and Multi-Spectral Method to Estimate Aerosol Optical Thickness over Land, for the Atmospheric Correction of FormoSat-2, LandSat, VEN μ S and Sentinel-2 Images. *Remote Sensing*, 7(3) :2668–2691. [46](#), [47](#), [118](#), [151](#)
- Hamrouni, Y., Paillassa, E., Chéret, V., Monteil, C., and Sheeren, D. (2021). From local to global : A transfer learning-based approach for mapping poplar plantations at national scale using Sentinel-2. *ISPRS Journal of Photogrammetry & Remote Sensing*, 171 :76–100. [170](#)

-
- Han, P., Chen, J., Han, Y., Yi, L., Zhang, Y., and Jiang, X. (2018). Monitoring rubber plantation distribution on Hainan Island using Landsat OLI imagery. *International Journal of Remote Sensing*, 39(8) :2189–2206. [25](#), [113](#)
- Hansen, J., Mitchard, E. A., and King, S. (2020). Assessing forest/non-forest separability using sentinel-1 c-band synthetic aperture radar. *Remote Sensing*, 12(11). [149](#)
- Hansen, M. C., Roy, D. P., Lindquist, E., Adusei, B., Justice, C. O., and Altstatt, A. (2008). A method for integrating MODIS and Landsat data for systematic monitoring of forest cover and change in the Congo Basin. *Remote Sensing of Environment*, 112(5) :2495–2513. [23](#), [24](#)
- Hastie, T., Tibshirani, R., and Friedman, J. (2009). Random Forests. In *The Elements of Statistical Learning*, pages 1–18. Springer New York, New York, NY. [63](#)
- Heilman, P. E. (1999). Planted forests : Poplars. In Boyle, J. R., Winjum, J. K., Kavanagh, K., and Jensen, E. C., editors, *Planted Forests : Contributions to the Quest for Sustainable Societies*, Forestry Sciences, pages 89–93. Springer Netherlands, Dordrecht. [3](#), [4](#)
- Henry, A. (1914). A new hybrid of poplar. *Gardener’s Chronicle, Series III.*, 56 :257–258. [4](#)
- Heyman, O., Gaston, G., Kimerling, A., and Campbell, J. (2003). A Per-Segment Approach to Improving Aspen Mapping from High-Resolution Remote Sensing Imagery. *Journal of Forestry*, 101(4) :29–33. [25](#), [113](#), [149](#)
- Hill, R. A., Wilson, A. K., George, M., and Hinsley, S. A. (2010). Mapping tree species in temperate deciduous woodland using time-series multi-spectral data. *Applied Vegetation Science*, 13(1) :86–99. [24](#)
- Hościło, A. and Lewandowska, A. (2019). Mapping Forest Type and Tree Species on a Regional Scale Using Multi-Temporal Sentinel-2 Data. *Remote Sensing*, 11(8) :929. [150](#)
- Hu, B., Miller, J. R., Zarco-Tejada, P., Freemantle, J., and Zwick, H. (2008). Boreal forest mapping at the BOREAS study area using seasonal optical indices sensitive to plant pigment content. *Canadian Journal of Remote Sensing*, 34(sup1) :S158–S171. [171](#)
- Hu, J., He, Z., Li, J., He, L., and Wang, Y. (2018). 3d-gabor inspired multiview active learning for spectral-spatial hyperspectral image classification. *Remote Sensing*, 10(7) :1070. [116](#)
- Hunt, E. and Rock, B. N. (1989). Detection of changes in leaf water content using near- and middle-infrared reflectances. *Remote Sensing of Environment*, 30(1) :43–54. [159](#)
- IFN (2011). Une nouvelle partition écologique et forestière du territoire métropolitain : Les sylvoécotésions (SER). Technical Report 26. [42](#)
- IGN (2013). Fiches descriptives des grandes régions écologiques (GRECO) et des sylvoécotésions (SER). Technical report. [41](#), [42](#)
- IGN (2014). Descriptif de contenu BD Forêt® Version 2.0. [10](#), [120](#)
- IGN (2018). Un inventaire forestier annuel sur l’ensemble de la France métropolitaine. Technical report, Service de l’inventaire forestier et environnemental. [11](#)
- IGN (2019). Le memento Inventaire Forestier. Technical report. [11](#)
- Immitzer, M., Atzberger, C., and Koukal, T. (2012). Tree Species Classification with Random Forest Using Very High Spatial Resolution 8-Band WorldView-2 Satellite Data. *Remote Sensing*, 4(9) :2661–2693. [149](#)



-
- Immitzer, M., Neuwirth, M., Böck, S., Brenner, H., Vuolo, F., and Atzberger, C. (2019). Optimal Input Features for Tree Species Classification in Central Europe Based on Multi-Temporal Sentinel-2 Data. *Remote Sensing*, 11(22) :2599. [24](#)
- Immitzer, M., Vuolo, F., and Atzberger, C. (2016). First Experience with Sentinel-2 Data for Crop and Tree Species Classifications in Central Europe. *Remote Sensing*, 8(3) :166. [24](#)
- Ines, A. V. M. and Honda, K. (2005). On quantifying agricultural and water management practices from low spatial resolution RS data using genetic algorithms : A numerical study for mixed-pixel environment. *Advances in Water Resources*, 28(8) :856–870. [29](#)
- Inglada, J. (2016). Otb Gapfilling, A Temporal Gapfilling For Image Time Series Library. [49](#), [50](#), [120](#)
- Inglada, J., Arias, M., Tardy, B., Hagolle, O., Valero, S., Morin, D., Dedieu, G., Sepulcre, G., Bontemps, S., Defourny, P., and Koetz, B. (2015). Assessment of an Operational System for Crop Type Map Production Using High Temporal and Spatial Resolution Satellite Optical Imagery. *Remote Sensing*, 7(9) :12356–12379. [49](#), [50](#), [122](#)
- Inglada, J., Vincent, A., Arias, M., and Marais-Sicre, C. (2016). Improved Early Crop Type Identification By Joint Use of High Temporal Resolution SAR And Optical Image Time Series. *Remote Sensing*, 8(5) :362. [122](#), [152](#)
- Inglada, J., Vincent, A., Arias, M., Tardy, B., Morin, D., and Rodes, I. (2017). Operational High Resolution Land Cover Map Production at the Country Scale Using Satellite Image Time Series. *Remote Sensing*, 9(1) :95. [32](#), [61](#), [160](#), [181](#)
- Iqbal, J. and Ali, M. (2020). Weakly-supervised domain adaptation for built-up region segmentation in aerial and satellite imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 167 :263–275. [34](#)
- Isebrands, J. G. and Richardson, J., editors (2013). *Poplars and Willows : Trees for Society and the Environment*. FAO, Rome. [113](#)
- Isebrands, J. G. and Richardson, J. (2014). *Poplars and Willows : Trees for Society and the Environment*. CABI. [3](#), [4](#)
- Jacquemoud, S., Verhoef, W., Baret, F., Bacour, C., Zarco-Tejada, P. J., Asner, G. P., François, C., and Ustin, S. L. (2009). PROSPECT+SAIL models : A review of use for vegetation characterization. *Remote Sensing of Environment*, 113 :S56–S66. [18](#)
- Jia, X., Kuo, B.-C., and Crawford, M. M. (2013). Feature mining for hyperspectral image classification. *Proceedings of the IEEE*, 101(3) :676–697. [154](#)
- Joly, D., Brossard, T., Cardot, H., Cavailhes, J., Hilal, M., and Wavresky, P. (2010). Les types de climats en france, une construction spatiale. *Cybergeo – European Journal of Geography*, 501. [151](#)
- Jones, T. G., Coops, N. C., and Sharma, T. (2010). Assessing the utility of airborne hyperspectral and LiDAR data for species distribution mapping in the coastal Pacific Northwest, Canada. *Remote Sensing of Environment*, 114(12) :2841–2852. [24](#)
- Joshi, A. J., Porikli, F., and Papanikolopoulos, N. (2009). Multi-class active learning for image classification. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2372–2379. [137](#), [140](#), [142](#)

-
- Jug, A., Hofmann-Schielle, C., Makeschin, F., and Rehfues, K. E. (1999). Short-rotation plantations of balsam poplars, aspen and willows on former arable land in the Federal Republic of Germany. II. Nutritional status and bioelement export by harvested shoot axes. *Forest Ecology and Management*, 121(1) :67–83. [6](#)
- Julien, Y. and Sobrino, J. A. (2010). Comparison of cloud-reconstruction methods for time series of composite NDVI data. *Remote Sensing of Environment*, 114(3) :618–625. [49](#), [122](#)
- Kahle, P., Baum, C., and Boelcke, B. (2005). Effect of afforestation on soil properties and mycorrhizal formation. *Pedosphere*, 15 :754–760. [6](#)
- Kandasamy, S., Baret, F., Verger, A., Neveux, P., and Weiss, M. (2013). A comparison of methods for smoothing and gap filling time series of remote sensing observations – application to MODIS LAI products. *Biogeosciences*, 10(6) :4055–4071. [49](#), [122](#)
- Kantorovich, L. V. (2006). On the Translocation of Masses. *Journal of Mathematical Sciences*, 133(4) :1381–1382. [70](#)
- Karasiak, N., Dejoux, J.-F., Fauvel, M., Willm, J., Monteil, C., and Sheeren, D. (2019). Statistical Stability and Spatial Instability in Mapping Forest Tree Species by Comparing 9 Years of Satellite Image Time Series. *Remote Sensing*, 11(21) :2512. [49](#), [52](#), [122](#), [150](#), [173](#), [200](#)
- Karasiak, N., Dejoux, J.-F., Monteil, C., and Sheeren, D. (2021). Spatial dependence between training and test sets : another pitfall of classification accuracy assessment in remote sensing. *Machine Learning*. [26](#), [156](#)
- Karasiak, N., Fauvel, M., Dejoux, J.-F., Monteil, C., and Sheeren, D. (2020). Optimal dates for deciduous tree species mapping using full years Sentinel-2 time series in south west France. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, V-3-2020 :469–476. [24](#)
- Kati Salovaara, Sirpa Thessler, Riffat Naseem Malik, and Hanna Tuomisto (2005). Classification of Amazonian primary rain forest vegetation using Landsat ETM+ satellite imagery. *Remote Sensing of Environment*, 97(1) :39–51. [23](#)
- Key, C. and Benson, N. (2006). Landscape Assessment : Ground measure of severity, the Composite Burn Index ; and Remote sensing of severity, the Normalized Burn Ratio. In *FIREMON : Fire Effects Monitoring and Inventory System*, pages LA 1–51. USDA Forest Service, Rocky Mountain Research Station. [159](#)
- Kim, D.-H., Sexton, J. O., Noojipady, P., Huang, C., Anand, A., Channan, S., Feng, M., and Townshend, J. R. (2014). Global, Landsat-based forest-cover change from 1990 to 2000. *Remote Sensing of Environment*, 155 :178–193. [23](#)
- Kim, M.-h., Madden, M., and Warner, T. (2009). Forest Type Mapping using Object-specific Texture Measures from Multispectral Ikonos Imagery : Segmentation Quality and Image Classification Issues. *undefined*. [24](#)
- Kivinen, S., Koivisto, E., Keski-Saari, S., Poikolainen, L., Tanhuanpää, T., Kuzmin, A., Viinikka, A., Heikkinen, R. K., Pykälä, J., Virkkala, R., Vihervaara, P., and Kumpula, T. (2020). A keystone species, European aspen (*Populus tremula* L.), in boreal forests : Ecological role, knowledge needs and mapping using remote sensing. *Forest Ecology and Management*, 462 :118008. [25](#), [149](#), [150](#)

-
- Lambert, J., Drenou, C., Denux, J.-P., Balent, G., and Cheret, V. (2013). Monitoring forest decline through remote sensing time series analysis. *GIScience & Remote Sensing*, 50(4) :437–457. [114](#)
- Lary, D. J., Alavi, A. H., Gandomi, A. H., and Walker, A. L. (2016). Machine learning in geosciences and remote sensing. *Geoscience Frontiers*, 7(1) :3–10. [56](#)
- Lausch, A., Erasmi, S., King, D., Magdon, P., and Heurich, M. (2016). Understanding Forest Health with Remote Sensing -Part I—A Review of Spectral Traits, Processes and Remote-Sensing Characteristics. *Remote Sensing*, 8(12) :1029. [22](#)
- Lazecky, M., Lhota, S., Penaz, T., and Klushina, D. (2018). Application of Sentinel-1 satellite to identify oil palm plantations in Balikpapan Bay. *IOP Conference Series : Earth and Environmental Science*, 169 :012064. [113](#)
- Le Bris, A., Chehata, N., Briottet, X., and Paparoditis, N. (2014). Use intermediate results of wrapper band selection methods : A first step toward the optimization of spectral configuration for land cover classifications. In *6th Workshop on Hyperspectral Image and Signal Processing : Evolution in Remote Sensing (WHISPERS)*, pages 1–4. [155](#)
- le Maire, G., Dupuy, S., Nouvellon, Y., Loos, R. A., and Hakamada, R. (2014). Mapping short-rotation plantations at regional scale using MODIS time series : Case of eucalypt plantations in Brazil. *Remote Sensing of Environment*, 152 :136–149. [25](#), [149](#)
- Le Toan, T., Quegan, S., Woodward, I., Lomas, M., Delbart, N., and Picard, G. (2004). Relating Radar Remote Sensing of Biomass to Modelling of Forest Carbon Budgets. *Climatic Change*, 67(2-3) :379–402. [57](#)
- Lefsky, M. A., Turner, D. P., Guzy, M., and Cohen, W. B. (2005). Combining lidar estimates of aboveground biomass and Landsat estimates of stand age for spatially extensive validation of modeled forest productivity. *Remote Sensing of Environment*, 95(4) :549–558. [22](#)
- Lehmann, E. A., Wallace, J. F., Caccetta, P. A., Furby, S. L., and Zdunic, K. (2013). Forest cover trends from time series Landsat data for the Australian continent. *International Journal of Applied Earth Observation and Geoinformation*, 21 :453–462. [23](#)
- Lewis, D. D. and Gale, W. A. (1994). A Sequential Algorithm for Training Text Classifiers. In Croft, B. W. and van Rijsbergen, C. J., editors, *SIGIR '94*, pages 3–12. Springer London, London. [124](#)
- Li, Z. and Fox, J. M. (2011). Integrating Mahalanobis typicalities with a neural network for rubber distribution mapping. *Remote Sensing Letters*, 2(2) :157–166. [25](#), [113](#)
- Li, Z. and Fox, J. M. (2012). Mapping rubber tree growth in mainland Southeast Asia using time-series MODIS 250 m NDVI and statistical data. *Applied Geography*, 32(2) :420–432. [25](#), [113](#), [149](#)
- Lier, P., Valorge, C., and Briottet, X. (2008). *Imagerie spatiale : des principes d'acquisition au traitement des images optiques pour l'observation de la terre*. Cépaduès, Toulouse, France. [15](#)
- Liu, P., Zhang, H., and Eom, K. (2017). Active deep learning for classification of hyperspectral images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(2) :712–724. [116](#)

-
- Liu, R., Shang, R., Liu, Y., and Lu, X. (2017a). Global evaluation of gap-filling approaches for seasonal NDVI with considering vegetation growth trajectory, protection of key point, noise resistance and curve stability. *Remote Sensing of Environment*, 189 :164–179. [49](#)
- Liu, X., Wang, Z., Shao, W., Ye, Z., and Zhang, J. (2017b). Phylogenetic and Taxonomic Status Analyses of the Abaso Section from Multiple Nuclear Genes and Plastid Fragments Reveal New Insights into the North America Origin of *Populus* (Salicaceae). *Frontiers in Plant Science*, 7. [3](#)
- Ma, L., Fu, T., and Li, M. (2018). Active learning for object-based image classification using predefined training objects. *International Journal of Remote Sensing*, 39(9) :2746–2765. [116](#)
- Main-Knorn, M., Cohen, W. B., Kennedy, R. E., Grodzki, W., Pflugmacher, D., Griffiths, P., and Hostert, P. (2013). Monitoring coniferous forest biomass change using a Landsat trajectory-based approach. *Remote Sensing of Environment*, 139 :277–290. [23](#)
- Malek, S., Miglietta, F., Gobakken, T., Næsset, E., Gianelle, D., and Dalponte, M. (2019). Optimizing Field Data Collection for Individual Tree Attribute Predictions Using Active Learning Methods. *Remote Sensing*, 11(8) :949. [116](#)
- Mallinis, G., Koutsias, N., Tsakiri-Strati, M., and Karteris, M. (2008). Object-based classification using Quickbird imagery for delineating forest vegetation polygons in a Mediterranean test site. *ISPRS Journal of Photogrammetry and Remote Sensing*, 63(2) :237–250. [24](#)
- Marie-Victorin, Brouillet, L., Rouleau, E., Goulet, I., and Hay, S. (2002). *Flore laurentienne*. G. Morin, Boucherville, Québec. OCLC : 49892585. [3](#)
- Marion Zapater, Benjamin Chapelet, and Lionel Coquelet (2012). Peupleraies et environnement : Un regard partagé sur la place de la peupleraie dans notre territoire. Brochure, CRPF Nord Pas-de-Calais Picardie et Centre Pour l’Agronomie et l’Agro-Industrie de la Province de Hainaut. [6](#)
- Markou, M. and Singh, S. (2003). Novelty detection : A review—part 1 : Statistical approaches. *Signal Processing*, 83(12) :2481–2497. [174](#)
- Matasci, G., Volpi, M., Kanevski, M., Bruzzone, L., and Tuia, D. (2015). Semisupervised Transfer Component Analysis for Domain Adaptation in Remote Sensing Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*, 53(7) :3550–3564. [115](#)
- McCallum, A. and Nigam, K. (1998). *Employing EM and Pool-Based Active Learning for Text Classification*. ICML '98. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA. [124](#)
- McIver, D. K. and Friedl, M. A. (2002). Using prior probabilities in decision-tree classification of remotely sensed data. *Remote Sensing of Environment*, 81(2) :253–261. [31](#), [32](#)
- McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, 12(2) :153–157. [156](#)
- Melaas, E. K., Friedl, M. A., and Zhu, Z. (2013). Detecting interannual variation in deciduous broadleaf forest phenology using Landsat TM/ETM+ data. *Remote Sensing of Environment*, 132 :176–185. [23](#)
- Mellor, A., Haywood, A., Stone, C., and Jones, S. (2013). The Performance of Random Forests in an Operational Setting for Large Area Sclerophyll Forest Classification. *Remote Sensing*, 5(6) :2838–2856. [61](#)



-
- Meroni, M., Rembold, F., Verstraete, M., Gommès, R., Schucknecht, A., and Beye, G. (2014). Investigating the Relationship between the Inter-Annual Variability of Satellite-Derived Vegetation Phenology and a Proxy of Biomass Production in the Sahel. *Remote Sensing*, 6(6) :5868–5884. [29](#)
- Meyer, H., Kühnlein, M., Appelhans, T., and Nauss, T. (2016). Comparison of four machine learning algorithms for their applicability in satellite-based optical rainfall retrievals. *Atmospheric Research*, 169 :424–433. [61](#)
- Mitra, P., Uma Shankar, B., and Pal, S. (2004). Segmentation of multispectral remote sensing images using active support vector machines. *Pattern Recognition Letters*, 25(9) :1067–1074. [116](#)
- Olofsson, P., Foody, G. M., Herold, M., Stehman, S. V., Woodcock, C. E., and Wulder, M. A. (2014). Good practices for estimating area and assessing accuracy of land change. *Remote Sensing of Environment*, 148 :42–57. [66](#), [67](#)
- Olthof, I., Butson, C., and Fraser, R. (2005). Signature extension through space for northern landcover classification : A comparison of radiometric correction methods. *Remote Sensing of Environment*, 95(3) :290–302. [32](#)
- Ørka, H. O., Dalponte, M., Gobakken, T., Næsset, E., and Ene, L. T. (2013). Characterizing forest species composition using multiple remote sensing data sources and inventory approaches. *Scandinavian Journal of Forest Research*, 28(7) :677–688. [22](#)
- Oshiro, T. M., Perez, P. S., and Baranauskas, J. A. (2012). How Many Trees in a Random Forest? In Hutchison, D., Kanade, T., Kittler, J., Kleinberg, J. M., Mattern, F., Mitchell, J. C., Naor, M., Nierstrasz, O., Pandu Rangan, C., Steffen, B., Sudan, M., Terzopoulos, D., Tygar, D., Vardi, M. Y., Weikum, G., and Perner, P., editors, *Machine Learning and Data Mining in Pattern Recognition*, volume 7376, pages 154–168. Springer Berlin Heidelberg, Berlin, Heidelberg. [63](#)
- Paillassa, É. (2014). Les peupleraies : Quels enjeux pour l’avenir de la populiculture française ? *Revue Forestière Française*, 66(3) :301–311. [8](#), [113](#)
- Pan, S. J. and Yang, Q. (2010). A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10) :1345–1359. [32](#), [33](#), [114](#), [115](#), [140](#)
- Pasolli, E., Melgani, F., Tuia, D., Pacifici, F., and Emery, W. J. (2014). Svm active learning approach for image classification using spatial information. *IEEE Transactions on Geoscience and Remote Sensing*, 52(4) :2217–2233. [116](#)
- Pasolli, E., Yang, H. L., and Crawford, M. M. (2016). Active-metric learning for classification of remotely sensed hyperspectral images. *IEEE Transactions on Geoscience and Remote Sensing*, 54(4) :1925–1939. [116](#)
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, É. (2011). Scikit-learn : Machine Learning in Python. *J. Mach. Learn. Res.*, 12 :2825–2830. [67](#), [85](#), [123](#), [157](#)
- Pelletier, C., Valero, S., Inglada, J., Champion, N., and Dedieu, G. (2016). Assessing the robustness of Random Forests to map land cover with high resolution satellite image time series over large areas. *Remote Sensing of Environment*, 187(Supplement C) :156–168. [61](#), [196](#)

-
- Pelletier, C., Valero, S., Inglada, J., Champion, N., Marais Sicre, C., and Dedieu, G. (2017). Effect of Training Class Label Noise on Classification Performances for Land Cover Mapping with Satellite Image Time Series. *Remote Sensing*, 9(2) :173. [61](#)
- Peñuelas, J., Gamon, J. A., Fredeen, A. L., Merino, J., and Field, C. B. (1994). Reflectance indices associated with physiological changes in nitrogen- and water-limited sunflower leaves. *Remote Sensing of Environment*, 48(2) :135–146. [159](#)
- Persello, C., Boularias, A., Dalponte, M., Gobakken, T., Næsset, E., and Schölkopf, B. (2014). Cost-sensitive active learning with lookahead : Optimizing field surveys for remote sensing data classification. *IEEE Transactions on Geoscience and Remote Sensing*, 52(10) :6652–6664. [116](#)
- Persello, C. and Bruzzone, L. (2012). Active learning for domain adaptation in the supervised classification of remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 50(11) :4468–4483. [116](#), [125](#)
- Persson, M., Lindberg, E., and Reese, H. (2018). Tree Species Classification with Multi-Temporal Sentinel-2 Data. *Remote Sensing*, 10(11) :1794. [24](#), [149](#)
- Petitjean, F., Ketterlin, A., and Gançarski, P. (2011). A Global Averaging Method for Dynamic Time Warping, with Applications to Clustering. *Pattern Recogn.*, 44(3) :678–693. [33](#), [115](#)
- Peyré, G. (2019a). Le transport optimal numérique et ses applications - Partie 1. [70](#)
- Peyré, G. (2019b). Le transport optimal numérique et ses applications - Partie 2. [73](#)
- Phua, M.-H. and Saito, H. (2003). Estimation of biomass of a mountainous tropical forest using Landsat TM data. *Canadian Journal of Remote Sensing*, 29(4) :429–440. [23](#)
- Poortinga, A., Tenneson, K., Shapiro, A., Nguyen, Q., San Aung, K., Chishtie, F., and Saah, D. (2019). Mapping Plantations in Myanmar by Fusing Landsat-8, Sentinel-2 and Sentinel-1 Data along with Systematic Error Quantification. *Remote Sensing*, 11(7) :831. [25](#), [113](#)
- Probst, P. and Boulesteix, A.-L. (2018). To Tune or Not to Tune the Number of Trees in Random Forest. *Journal of Machine Learning Research*, page 18. [63](#)
- Pudil, P., Novovičová, J., and Kittler, J. (1994). Floating search methods in feature selection. *Pattern Recognition Letters*, 15(11) :1119–1125. [155](#)
- Puttinaovaratt, S. and Horkaew, P. (2019). Deep and machine learnings of remotely sensed imagery and its multi-band visual features for detecting oil palm plantation. *Earth Science Informatics*, 12(4) :429–446. [25](#), [149](#)
- Rajan, S., Ghosh, J., and Crawford, M. M. (2008). An Active Learning Approach to Hyperspectral Data Classification. *IEEE Transactions on Geoscience and Remote Sensing*, 46(4) :1231–1242. [34](#), [116](#), [137](#)
- Raschka, S. (2018). Mlxtend : Providing machine learning and data science utilities and extensions to python’s scientific computing stack. *The Journal of Open Source Software*, 3(24). [157](#)
- Reed, B. C., Brown, J. F., VanderZee, D., Loveland, T. R., Merchant, J. W., and Ohlen, D. O. (1994). Measuring phenological variability from satellite imagery. *Journal of Vegetation Science*, 5(5) :703–714. [29](#)

-
- Reitberger, J., Krzystek, P., and Stilla, U. (2008). Analysis of full waveform LIDAR data for the classification of deciduous and coniferous trees. *International Journal of Remote Sensing*, 29(5) :1407–1431. [149](#)
- Richards, J. A. (2013). Supervised Classification Techniques. In *Remote Sensing Digital Image Analysis : An Introduction*, pages 247–318. Springer Berlin Heidelberg, Berlin, Heidelberg. [58](#)
- Richardson, A. D., Duigan, S. P., and Berlyn, G. P. (2002). An evaluation of noninvasive methods to estimate foliar chlorophyll content. *New phytologist*, 153(1) :185–194. [245](#)
- Rikimaru, A., Roy, P., and Miyatake, S. (2002). Tropical forest cover density mapping. *Tropical ecology*, 43(1) :39–47. [245](#)
- Robert, A. (2018). Poplar plantations in France, at the heart of a conflict between provisioning services and cultural (dis)services. In *Ecosystem Services in a Changing World : Moving from Theory to Practice, San Sebastián, Spain*. [113](#)
- Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., and Rigol-Sanchez, J. P. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 67 :93–104. [62](#)
- Rondeaux, G., Steven, M., and Baret, F. (1996). Optimization of soil-adjusted vegetation indices. *Remote Sensing of Environment*, 55(2) :95–107. [159](#)
- Rosenqvist, A. (1996). Evaluation of JERS-1, ERS-1 and Almaz SAR backscatter for rubber and oil palm stands in West Malaysia. *International Journal of Remote Sensing - INT J REMOTE SENS*, 17 :3219–3231. [113](#)
- Rouse, J., Haas, R. H., Schell, J. A., and Deering, D. (1974). Monitoring vegetation systems in the great plains with erts. In *Third earth resources technology satellite-1 symposium, Vol. 1, 10-14 December, 1973*, pages 309–317. NASA Scientific and Technical Information Office Washington, DC. [159](#)
- Rüetschi, M., Schaepman, M. E., and Small, D. (2018). Using Multitemporal Sentinel-1 C-band Backscatter to Monitor Phenology and Classify Deciduous and Coniferous Forests in Northern Switzerland. *Remote Sensing*, 10(1) :55. [24](#)
- S. Garea, A., Heras, D. B., and Argüello, F. (2019). TCANet for Domain Adaptation of Hyperspectral Images. *Remote Sensing*, 11(19) :2289. [34](#)
- Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, 3(3) :210–229. [56](#)
- Sarsekova, D. (2015). Growth and Productivity of Poplar Species in Southeastern Kazakhstan. *Universal Journal of Plant Science*, 3(6) :132–140. [173](#)
- Segal, M. R. (2004). Machine Learning Benchmarks and Random Forest Regression. *UCSF : Center for Bioinformatics and Molecular Biostatistics*. [62](#), [63](#)
- SEINet Portal Network (2021). Species with Images. [http:// :swbiodiversity.org/seinet/index.php](http://swbiodiversity.org/seinet/index.php). [5](#)
- Settles, B. (2012). Active Learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 6(1) :1–114. [124](#), [125](#)

-
- Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, 27(3) :379–423. [126](#)
- Sheeren, D., Fauvel, M., Josipović, V., Lopes, M., Planque, C., Willm, J., and Dejoux, J.-F. (2016). Tree Species Classification in Temperate Forests Using Formosat-2 Satellite Image Time Series. *Remote Sensing*, 8(9) :734. [26](#), [29](#), [149](#)
- Shimada, M., Itoh, T., Motooka, T., Watanabe, M., Shiraishi, T., Thapa, R., and Lucas, R. (2014). New global forest/non-forest maps from ALOS PALSAR data (2007–2010). *Remote Sensing of Environment*, 155 :13–31. [24](#)
- Spracklen, B. and Spracklen, D. V. (2021). Synergistic use of sentinel-1 and sentinel-2 to map natural forest and acacia plantation and stand ages in north-central vietnam. *Remote Sensing*, 13(2). [25](#), [149](#)
- Spracklen, B. D. and Spracklen, D. V. (2019). Identifying European Old-Growth Forests using Remote Sensing : A Study in the Ukrainian Carpathians. *Forests*, 10(2) :127. [22](#)
- Springer, K. R., Wang, R., and Gamon, J. A. (2017). Parallel Seasonal Patterns of Photosynthesis, Fluorescence, and Reflectance Indices in Boreal Trees. *Remote Sensing*, 9(7) :691. [172](#)
- Stettler, R. F. (2009). Cottonwood and the River of Time. [4](#)
- Stobrawa, K. (2014). Poplars (*Populus* spp.) : Ecological Role, Applications and Scientific Perspectives in the 21st Century (Review paper). *Baltic Forestry*, 20 :204–213. [6](#)
- Stoffels, J., Hill, J., Sachtleber, T., Mader, S., Buddenbaum, H., Stern, O., Langshausen, J., Dietz, J., and Ontrup, G. (2015). Satellite-Based Derivation of High-Resolution Forest Information Layers for Operational Forest Management. *Forests*, 6(12) :1982–2013. [24](#)
- Stout, A. B. and Schreiner, E. J. (1933). Results of a project in hybridizing poplars. *Journal of Heredity*, 24(2) :16–229. [4](#)
- Stumpf, A., Lachiche, N., Malet, J., Kerle, N., and Puissant, A. (2014). Active Learning in the Spatial Domain for Remote Sensing Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*, 52(5) :2492–2507. [116](#)
- Sun, Y., Qin, Q., Ren, H., Zhang, T., and Chen, S. (2020). Red-Edge Band Vegetation Indices for Leaf Area Index Estimation From Sentinel-2/MSI Imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 58(2) :826–840. [171](#)
- Tardy, B., Inglada, J., and Michel, J. (2019). Assessment of Optimal Transport for Operational Land-Cover Mapping Using High-Resolution Satellite Images Time Series without Reference Data of the Mapping Period. *Remote Sensing*, 11(9) :1047. [34](#), [81](#), [108](#)
- Tong, X.-Y., Xia, G.-S., Lu, Q., Shen, H., Li, S., You, S., and Zhang, L. (2020). Land-cover classification with high-resolution remote sensing images using transferable deep models. *Remote Sensing of Environment*, 237 :111322. [34](#)
- Torbick, N., Ledoux, L., Salas, W., and Zhao, M. (2016). Regional mapping of plantation extent using multisensor imagery. *Remote Sensing*, 8(3). [25](#), [149](#)
- Tou, J. T., Tou, T.-C., and Gonzalez, R. C. (1974). *Pattern Recognition Principles*. Addison-Wesley Publishing Company. [57](#)

-
- Tucker, C. J., Justice, C. O., and Prince, S. D. (1986). Monitoring the grasslands of the Sahel 1984-1985. *International Journal of Remote Sensing*, 7(11) :1571–1581. [171](#)
- Tuia, D., Pasolli, E., and Emery, W. J. (2011a). Using active learning to adapt remote sensing image classifiers. *Remote Sensing of Environment*, 115(9) :2232–2242. [115](#), [116](#), [140](#), [142](#)
- Tuia, D., Persello, C., and Bruzzone, L. (2016). Domain Adaptation for the Classification of Remote Sensing Data : An Overview of Recent Advances. *IEEE Geoscience and Remote Sensing Magazine*, 4(2) :41–57. [30](#), [32](#), [33](#), [115](#)
- Tuia, D., Ratle, F., Pacifici, F., Kanevski, M., and Emery, W. (2009). Active Learning Methods for Remote Sensing Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*, 47(7) :2218–2232. [114](#)
- Tuia, D., Volpi, M., Copa, L., Kanevski, M., and Munoz-Mari, J. (2011b). A Survey of Active Learning Algorithms for Supervised Remote Sensing Image Classification. *IEEE Journal of Selected Topics in Signal Processing*, 5(3) :606–617. [115](#), [116](#), [124](#)
- Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind*, LIX(236) :433–460. [56](#)
- Ustin, S. L. and Middleton, E. M. (2021). Current and near-term advances in Earth observation for ecological applications. *Ecological Processes*, 10(1). [21](#)
- Valovcin, F. R. (1976). *Snow/cloud discrimination*. Number 349. Air Force Geophysics Laboratories, Air Force Systems Command, United States [245](#)
- Vangi, E., D’Amico, G., Francini, S., Giannetti, F., Lasserre, B., Marchetti, M., and Chirici, G. (2021). The New Hyperspectral Satellite PRISMA : Imagery for Forest Types Discrimination. *Sensors*, 21(4) :1182. [21](#)
- Viinikka, A., Hurskainen, P., Keski-Saari, S., Kivinen, S., Tanhuanpää, T., Mäyrä, J., Poikolainen, L., Vihervaara, P., and Kumpula, T. (2020). Detecting European Aspen (*Populus tremula* L.) in Boreal Forests Using Airborne Hyperspectral and Airborne Laser Scanning Data. *Remote Sensing*, 12(16) :2610. [25](#), [26](#), [149](#), [171](#)
- Villar, M. and Forestier, O. (2017). La france à la sauvegarde du peuplier noir : état actuel du programme de conservation et de valorisation des ressources génétiques. *Revue Forestière Française*, LXIX(3) :195–204. [173](#)
- Vlachos, A. (2008). A stopping criterion for active learning. *Computer Speech & Language*, 22(3) :295–312. [124](#)
- Volpi, M., Tuia, D., and Kanevski, M. (2012). Memory-based cluster sampling for remote sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 50(8) :3096–3106. [116](#)
- Walker, J., de Beurs, K., Wynne, R., and Gao, F. (2012). Evaluation of Landsat and MODIS data fusion products for analysis of dryland forest phenology. *Remote Sensing of Environment*, 117 :381–393. [23](#), [24](#)
- Wemmert, C., Puissant, A., Forestier, G., and Gancarski, P. (2009). Multiresolution Remote Sensing Image Clustering. *IEEE Geoscience and Remote Sensing Letters*, 6(3) :533–537. [57](#)
- Wilson, G. and Cook, D. J. (2020). A Survey of Unsupervised Deep Domain Adaptation. *arXiv :1812.02849 [cs, stat]*. [34](#)

-
- Woodcock, C. E., Collins, J. B., Gopal, S., Jakabhazy, V. D., Li, X., Macomber, S., Ryherd, S., Judson Harward, V., Levitan, J., Wu, Y., and Warbington, R. (1994). Mapping forest vegetation using Landsat TM imagery and a canopy reflectance model. *Remote Sensing of Environment*, 50(3) :240–254. [23](#)
- Woodcock, C. E., Macomber, S. A., Pax-Lenney, M., and Cohen, W. B. (2001). Monitoring large areas for forest change using Landsat : Generalization across space, time and Landsat sensors. *Remote Sensing of Environment*, 78(1) :194–203. [32](#), [114](#)
- Wulder, M. A. and Franklin, S. E. (2003). Remote Sensing of Forest Environments, Introduction. In Wulder, M. A. and Franklin, S. E., editors, *Remote Sensing of Forest Environments : Concepts and Case Studies*, pages 3–12. Springer US, Boston, MA. [13](#)
- Wulder, M. A., Masek, J. G., Cohen, W. B., Loveland, T. R., and Woodcock, C. E. (2012). Opening the archive : How free data has enabled the science and monitoring promise of Landsat. *Remote Sensing of Environment*, 122 :2–10. [28](#)
- Xiao, C., Li, P., and Feng, Z. (2019). Monitoring annual dynamics of mature rubber plantations in Xishuangbanna during 1987-2018 using Landsat time series data : A multiple normalization approach. *International Journal of Applied Earth Observation and Geoinformation*, 77 :30–41. [25](#), [113](#), [149](#)
- Ye, S., Rogan, J., and Sangermano, F. (2018). Monitoring rubber plantation expansion using Landsat data time series and a Shapelet-based approach. *ISPRS Journal of Photogrammetry and Remote Sensing*, 136 :134–143. [25](#), [113](#), [149](#)
- Yin, H., Brandão, A., Buchner, J., Helmers, D., Iuliano, B. G., Kimambo, N. E., Lewińska, K. E., Razenkova, E., Rizayeva, A., Rogova, N., Spawn, S. A., Xie, Y., and Radeloff, V. C. (2020). Monitoring cropland abandonment with Landsat time series. *Remote Sensing of Environment*, 246 :111873. [32](#)
- Zarco-Tejada, P. J. and Miller, J. R. (1999). Land cover mapping at BOREAS using red edge spectral parameters from CASI imagery. *Journal of Geophysical Research : Atmospheres*, 104(D22) :27921–27933. [171](#)
- Zhang, F., Xie, C., Li, K., Xu, M., Wang, X., and Xia, Z. (2012). Forest and deforestation identification based on multitemporal polarimetric RADARSAT-2 images in Southwestern China. *Journal of Applied Remote Sensing*, 6(1) :063527. [24](#)
- Zhang, H., Morison, J. I. L., and Simmonds, L. P. (1999). Transpiration and water relations of poplar trees growing close to the water table. *Tree Physiology*, 19(9) :563–573. [171](#)
- Zhang, Z., Pasolli, E., and Crawford, M. (2020). An adaptive multiview active learning approach for spectral–spatial classification of hyperspectral images. *IEEE Transactions on Geoscience and Remote Sensing*, 58(4) :2557–2570. [116](#)
- Zhang, Z., Pasolli, E., Yang, H., and Crawford, M. (2016). Multimetric active learning for classification of remote sensing data. *IEEE Geoscience and Remote Sensing Letters*, 13(7) :1007–1011. [116](#)
- Zhong, P., Gong, Z., Li, S., and Schönlieb, C.-B. (2017). Learning to Diversify Deep Belief Networks for Hyperspectral Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*, 55(6) :3516–3530. [125](#)

Annexes

Annexe A

Compléments chapitre 5

A.1 Acquisition dates of the Sentinel-2 images

Table A.1 – Acquisition dates in each Sentinel-2 tile over 2017.

| 31UEQ (north-east) | 30TYT (centre) | 31TCJ (south-west) | Resampled dates (common to all three tiles) |
|-------------------------------------|---------------------------------|-------------------------------------|--|
| 16/01/2017 | 19/01/2017 | 26/01/2017 | 26/01/2017 |
| 26/01/2017 | 29/01/2017 | 15/02/2017 | 05/02/2017 |
| 15/02/2017 | 08/02/2017 | 25/02/2017 | 15/02/2017 |
| 25/02/2017 | 18/02/2017 | 17/03/2017 | 25/02/2017 |
| 27/03/2017 | 10/03/2017 | 27/03/2017 | 07/03/2017 |
| 26/04/2017 | 30/03/2017 | 06/04/2017 | 17/03/2017 |
| 16/05/2017 | 09/04/2017 | 06/05/2017 | 27/03/2017 |
| 26/05/2017 | 19/04/2017 | 16/05/2017 | 06/04/2017 |
| 05/06/2017 | 29/04/2017 | 26/05/2017 | 16/04/2017 |
| 15/06/2017 | 09/05/2017 | 05/06/2017 | 26/04/2017 |
| 25/06/2017 | 19/05/2017 | 25/06/2017 | 06/05/2017 |
| 05/07/2017 | 29/05/2017 | 05/07/2017 | 16/05/2017 |
| 25/07/2017 | 18/06/2017 | 04/08/2017 | 26/05/2017 |
| 04/08/2017 | 28/06/2017 | 14/08/2017 | 05/06/2017 |
| 14/08/2017 | 18/07/2017 | 19/08/2017 | 15/06/2017 |
| 24/08/2017 | 28/07/2017 | 24/08/2017 | 25/06/2017 |
| 29/08/2017 | 02/08/2017 | 29/08/2017 | 05/07/2017 |
| 03/09/2017 | 07/08/2017 | 03/09/2017 | 15/07/2017 |
| 23/09/2017 | 22/08/2017 | 08/09/2017 | 25/07/2017 |
| 03/10/2017 | 27/08/2017 | 13/09/2017 | 04/08/2017 |
| 13/10/2017 | 06/09/2017 | 23/09/2017 | 14/08/2017 |
| 18/10/2017 | 11/09/2017 | 28/09/2017 | 24/08/2017 |
| 07/11/2017 | 16/09/2017 | 03/10/2017 | 03/09/2017 |
| 22/11/2017 | 21/09/2017 | 08/10/2017 | 13/09/2017 |
| 02/12/2017 | 06/10/2017 | 13/10/2017 | 23/09/2017 |
| 12/12/2017 | 11/10/2017 | 18/10/2017 | 03/10/2017 |
| | 16/10/2017 | 23/10/2017 | 13/10/2017 |
| | 21/10/2017 | 28/10/2017 | 23/10/2017 |
| | 26/10/2017 | 02/11/2017 | 02/11/2017 |
| | 31/10/2017 | 07/11/2017 | 12/11/2017 |
| | 05/11/2017 | 17/11/2017 | 22/11/2017 |
| | 15/11/2017 | 22/11/2017 | 02/12/2017 |
| | 25/11/2017 | 27/11/2017 | 12/12/2017 |
| | 15/12/2017 | 07/12/2017 | 22/12/2017 |
| | | 12/12/2017 | |
| | | 27/12/2017 | |

A.2 Example of the gap-filling process

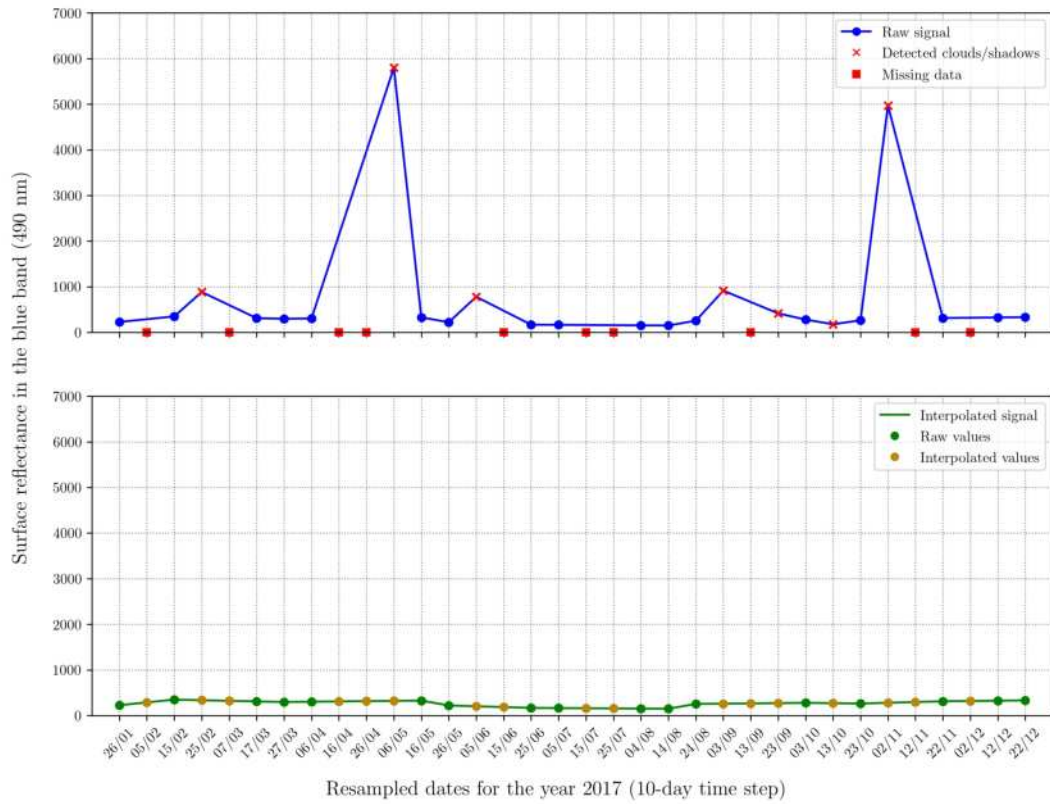


Figure A.1 – Example of the gap-filling process for a poplar pixel extracted from the Sentinel-2 time series for the year 2017 (31TCJ tile).

A.3 Classification results before the Active Learning adaptation

A.3.1 Source tile: north-eastern (31UEQ)

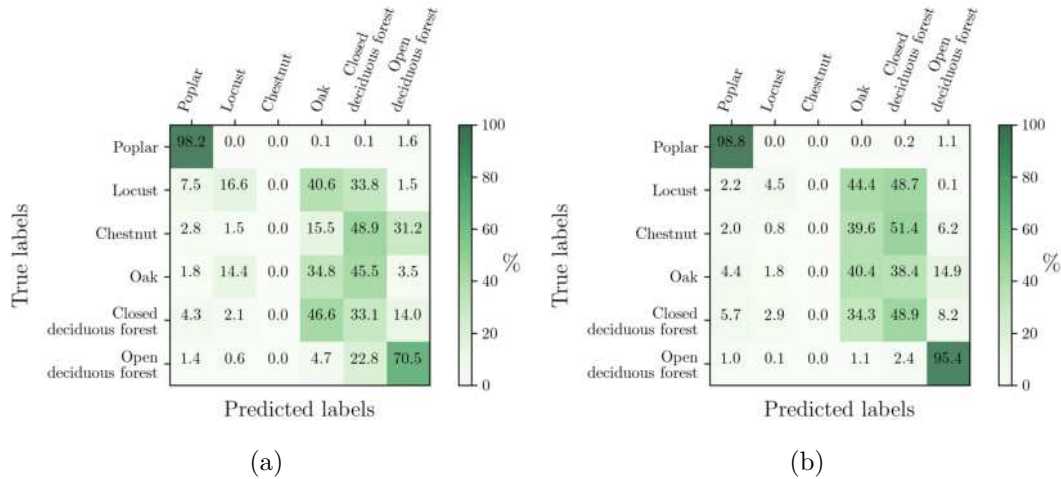


Figure A.2 – Normalised confusion matrices obtained before active learning adaptation (before adding new samples from target) when the transfer was performed from the north-eastern source tile (31UEQ) to (a) the central (30TYT) and (b) the south-western (31TCJ) target tiles.

A.3.2 Source tile: central (30TYT)

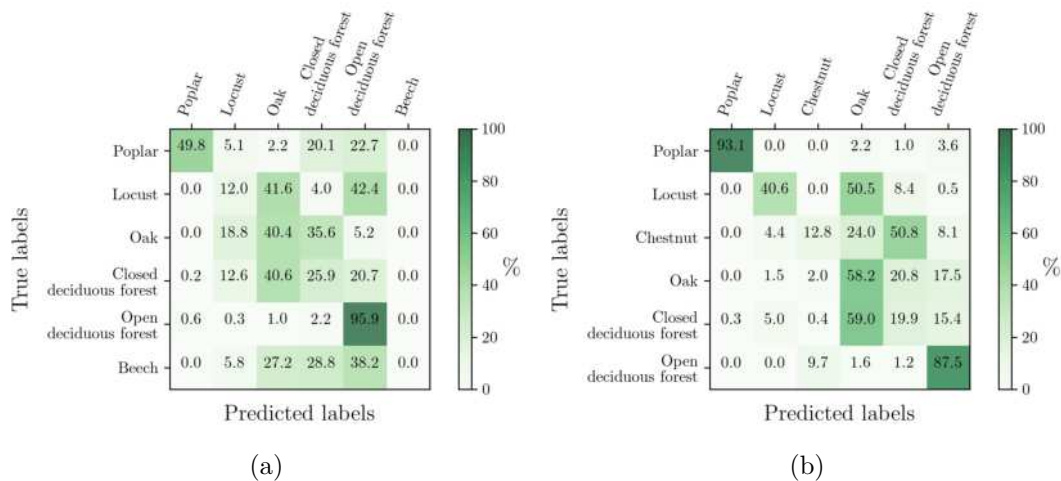


Figure A.3 – Normalised confusion matrices obtained before active learning adaptation (before adding new samples from target) when the transfer was performed from the central source tile (30TYT) to (a) the north-eastern (31UEQ) and (b) the south-western (31TCJ) target tiles.

A.3.3 Source tile: south-western (31TCJ)

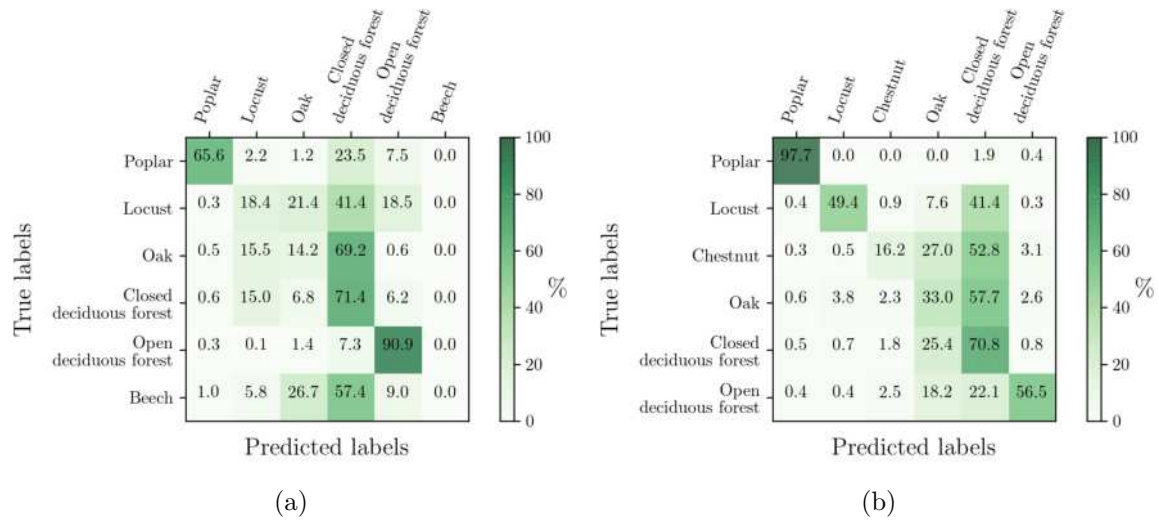
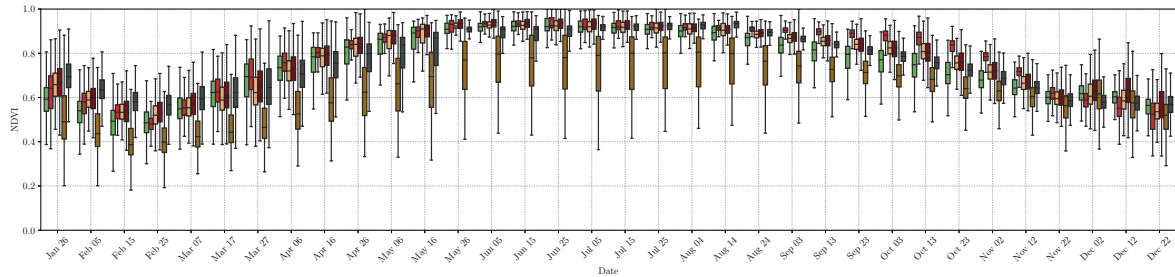
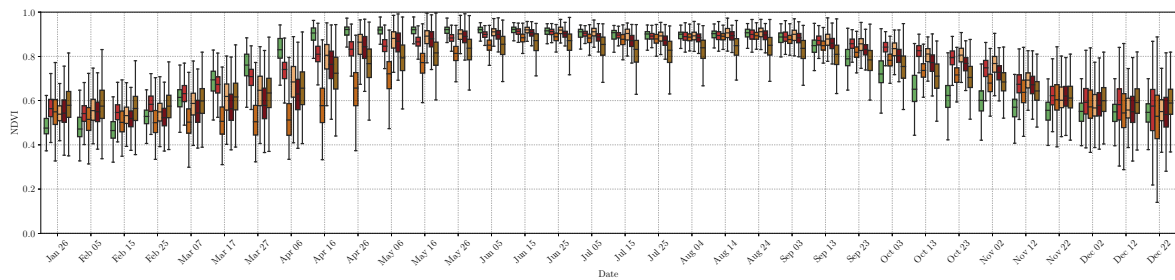


Figure A.4 – Normalised confusion matrices obtained before active learning adaptation (before adding new samples from target) when the transfer was performed from the south-western source tile (31TCJ) to (a) the north-eastern (31UEQ) and (b) the central (30TYT) target tiles.

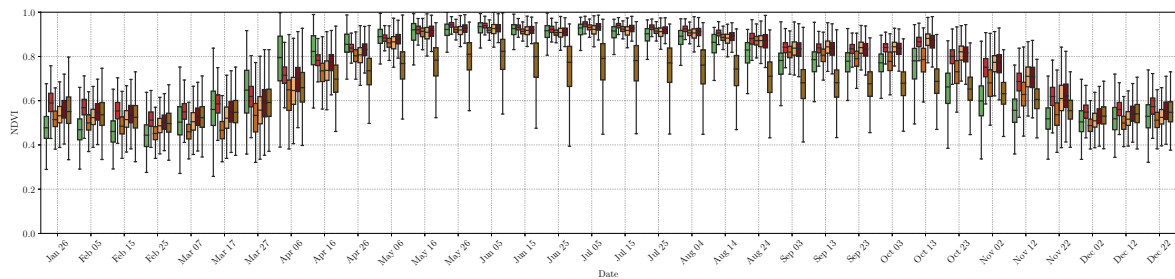
A.4 Distribution of NDVI values within the deciduous classes of each tile



(a)



(b)



(c)

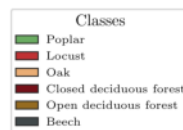


Figure A.5 – Spectral variability of Sentinel-2 NDVI over the deciduous classes in (a) the north-eastern (31UEQ), (b) the central (30TYT) and (c) the south-western (31TCJ) tiles.

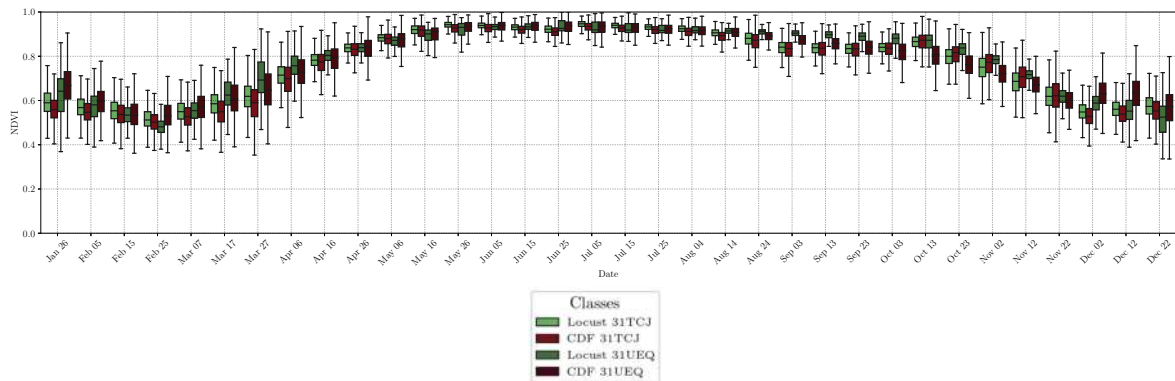


Figure A.6 – Spectral variability of Sentinel-2 NDVI between locusts and closed deciduous forests (CDF) on the north-eastern (N.E) and south-western (S.W) tiles.

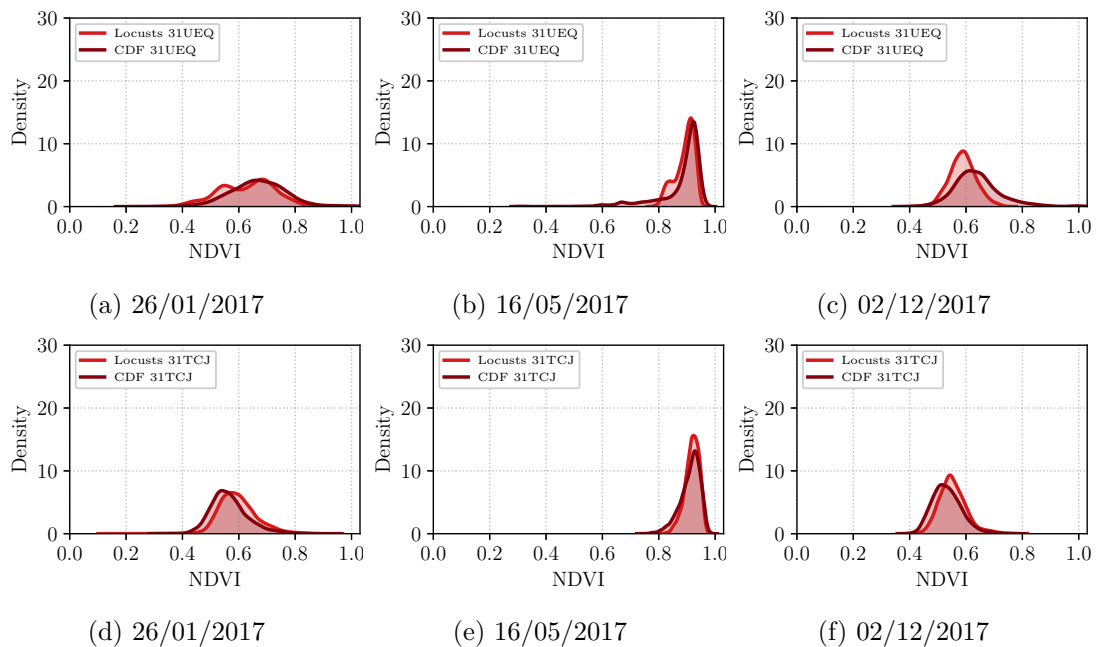
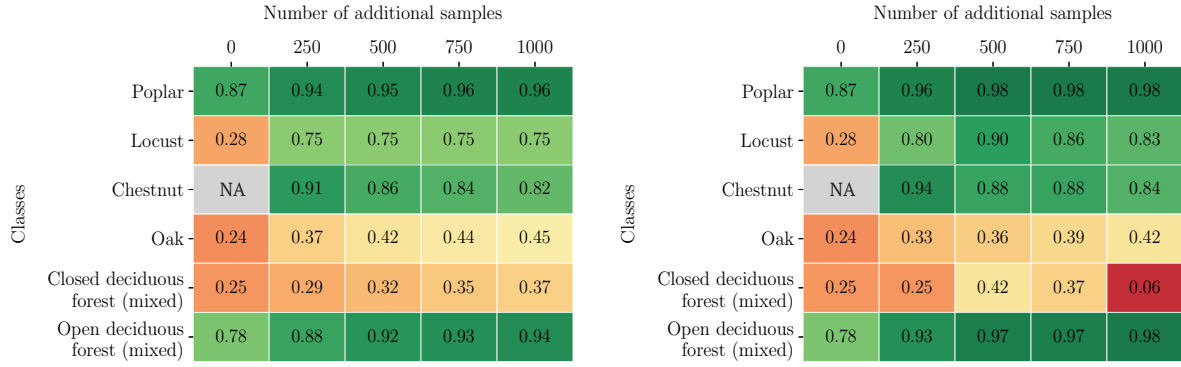


Figure A.7 – Kernel density of NDVI for locusts (red line) and closed deciduous forest (CDF) (burgundy line) in the north-eastern (31UEQ) and south-western (31TCJ) tiles for January 26th (a) and (d), May 16th (b) and (e) and December 2nd (c) and (f).

A.5 Active learning vs. random sampling: class-specific assessment according to producer's and user's accuracy measures

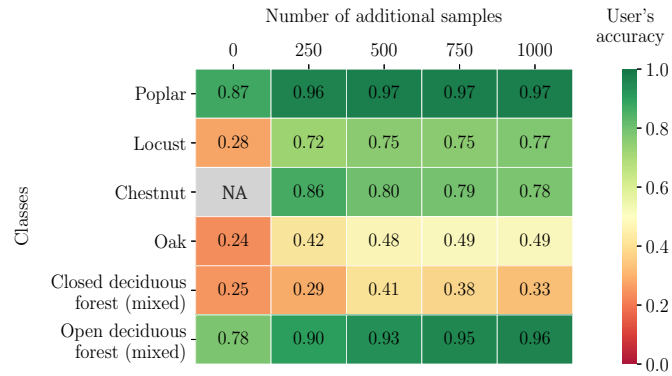


Figure A.8 – Average producer's accuracy per class obtained over 30 independent repetitions as a function of the number of added samples with (a) random sampling, (b) entropy-based active learning (AL_H) and (c) margin sampling-based active learning (AL_{MS}). The transfer was performed from the north-eastern tile (31UEQ) to the south-western tile (31TCJ).



(a) Random selection (RS)

(b) Active learning (AL_H)



(c) Active learning (AL_{MS})

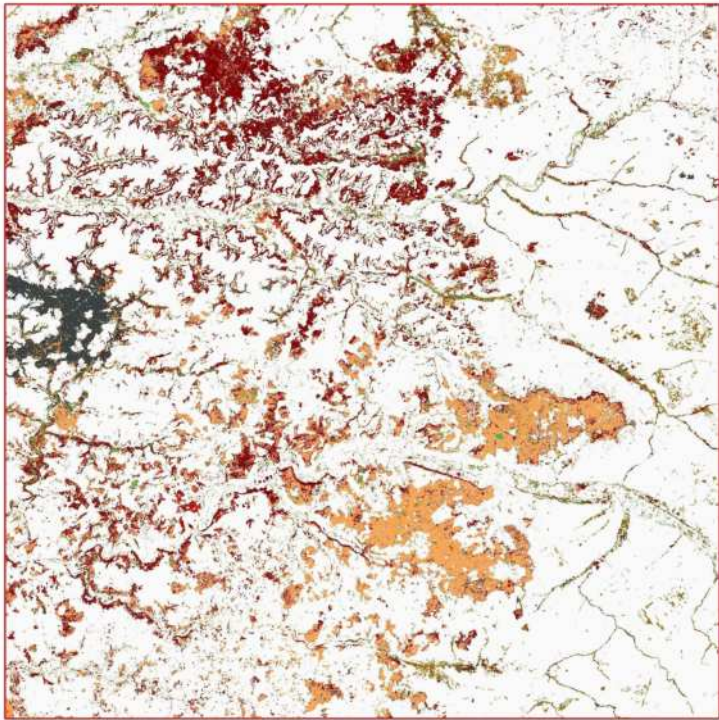
Figure A.9 – Average user's accuracy per class obtained over 30 independent repetitions as a function of the number of added samples with (a) random sampling, (b) entropy-based active learning (AL_H) and (c) margin sampling-based active learning (AL_{MS}). The transfer was performed from the north-eastern tile (31UEQ) to the south-western tile (31TCJ).

A.6 Active learning results with a combination of uncertainty and diversity measures

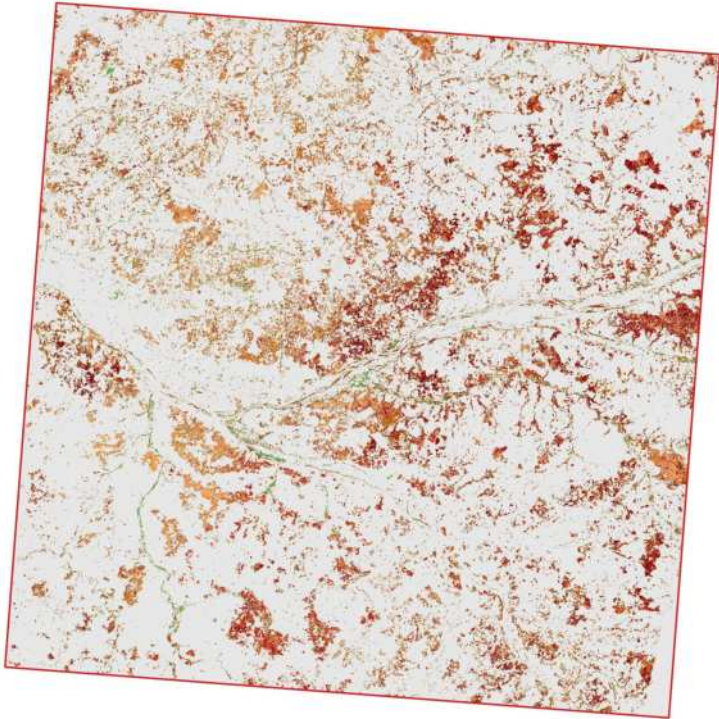
Table A.2 – Overall accuracy values (in %) according to the direction of the transfer and the number of target samples added to the initial set for active learning with margin sampling (MS) uncertainty metric combined with Euclidean distance (AL_{MS-Eu}) and Cosine (AL_{MS-C}) diversity metrics. The values are averaged over 30 independent repetitions.

| Source tiles | OA (%) | Target tiles | | | | | | | | | | | | | | |
|--------------|--------------|---------------------------|-----|-----|-----|------|-------|-----|-----|-----|------|-------|-----|-----|-----|------|
| | | 31UEQ | | | | | 30TYT | | | | | 31TCJ | | | | |
| | | No. of additional samples | | | | | | | | | | | | | | |
| | | 0 | 250 | 500 | 750 | 1000 | 0 | 250 | 500 | 750 | 1000 | 0 | 250 | 500 | 750 | 1000 |
| 31UEQ | AL_{MS-Eu} | - | - | - | - | - | 36 | 53 | 59 | 62 | 63 | 47 | 62 | 69 | 71 | 73 |
| | AL_{MS-C} | - | - | - | - | - | 36 | 50 | 57 | 60 | 62 | 47 | 55 | 64 | 68 | 70 |
| 30TYT | AL_{MS-Eu} | 31 | 47 | 53 | 56 | 57 | - | - | - | - | - | 53 | 66 | 69 | 70 | 71 |
| | AL_{MS-C} | 31 | 46 | 52 | 55 | 57 | - | - | - | - | - | 53 | 64 | 67 | 69 | 70 |
| 31TCJ | AL_{MS-Eu} | 40 | 53 | 61 | 63 | 64 | 52 | 62 | 64 | 65 | 65 | - | - | - | - | - |
| | AL_{MS-C} | 40 | 54 | 60 | 63 | 64 | 52 | 61 | 64 | 65 | 65 | - | - | - | - | - |

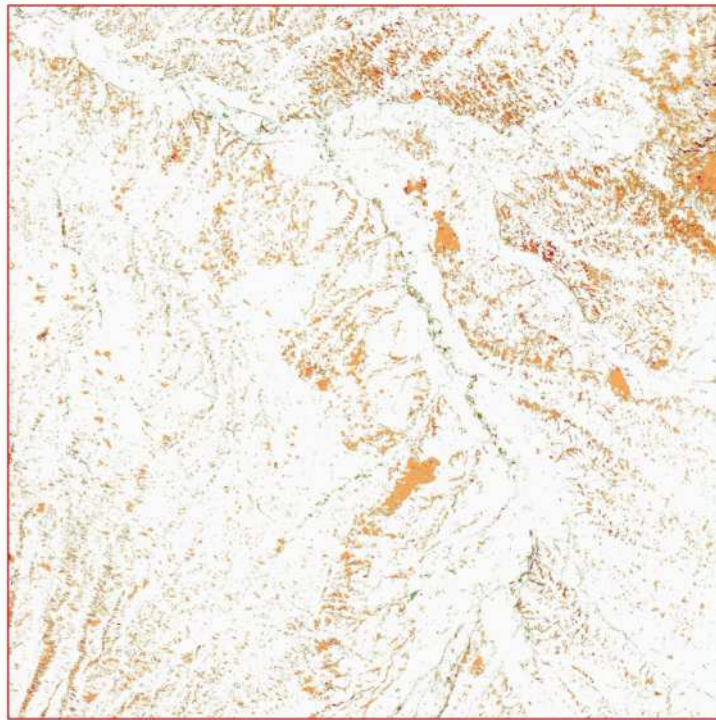
A.7 Complete prediction maps



(a) North-eastern tile 31UEQ



(b) Central tile 30TYT



(c) South-western tile 31TCJ

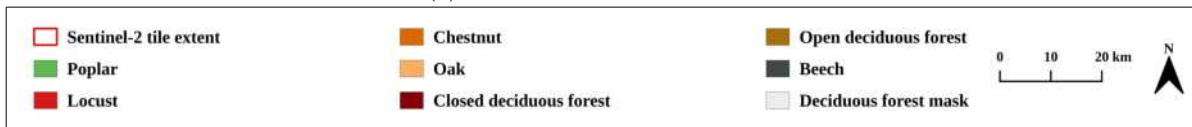


Figure A.10 – Prediction maps of the (a) north-eastern (31UEQ), (b) central (30TYT) and (c) south-western (31TCJ) tiles according to the three-tile global model built by AL.

Annexe B

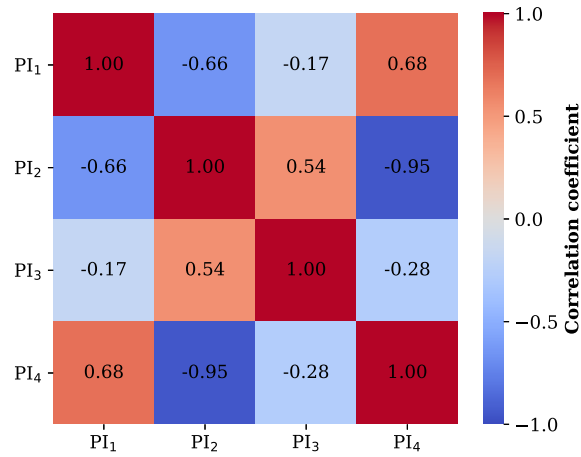
Compléments chapitre 6

B.1 Resampled dates for 2017 and 2018 time series

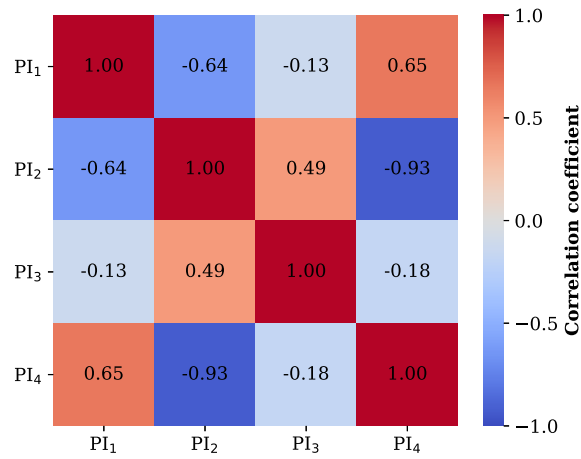
Table B.1 – Common resampled dates over 2017 and 2018 time series.

| Resampled dates (in 2017) | Resampled dates (in 2018) |
|--------------------------------------|--------------------------------------|
| NA | 6/1/2018 |
| NA | 16/1/2018 |
| 26/1/2017 | 26/1/2018 |
| 5/2/2017 | 5/2/2018 |
| 15/2/2017 | 15/2/2018 |
| 25/2/2017 | 25/2/2018 |
| 7/3/2017 | 7/3/2018 |
| 17/3/2017 | 17/3/2018 |
| 27/3/2017 | 27/3/2018 |
| 6/4/2017 | 6/4/2018 |
| 16/4/2017 | 16/4/2018 |
| 26/4/2017 | 26/4/2018 |
| 6/5/2017 | 6/5/2018 |
| 16/5/2017 | 16/5/2018 |
| 26/5/2017 | 26/5/2018 |
| 5/6/2017 | 5/6/2018 |
| 15/6/2017 | 15/6/2018 |
| 25/6/2017 | 25/6/2018 |
| 5/7/2017 | 5/7/2018 |
| 15/7/2017 | 15/7/2018 |
| 25/7/2017 | 25/7/2018 |
| 4/8/2017 | 4/8/2018 |
| 14/8/2017 | 14/8/2018 |
| 24/8/2017 | 24/8/2018 |
| 3/9/2017 | 3/9/2018 |
| 13/9/2017 | 13/9/2018 |
| 23/9/2017 | 23/9/2018 |
| 3/1/2017 | 3/1/2018 |
| 13/1/2017 | 13/1/2018 |
| 23/1/2017 | 23/1/2018 |
| 2/11/2017 | 2/11/2018 |
| 12/11/2017 | 12/11/2018 |
| 22/11/2017 | 22/11/2018 |
| 2/12/2017 | 2/12/2018 |
| 12/12/2017 | 12/12/2018 |
| 22/12/2017 | 22/12/2018 |

B.2 Correlation between the poplar index (PI) variants



(a) year 2017



(b) year 2018

Figure B.1 – Spearman’s ρ correlation coefficients between poplar index (PI) variants based on all resampled dates for 2017 and 2018 S2 time series.

B.3 Additional spectral indices tested

Table B.2 – List of additional spectral indices tested and their corresponding formulas.

| Indices | Formulas | References |
|--------------------------------------|--|---------------------------|
| Normalised Difference Moisture Index | $NDMI = \frac{B8-B11}{B8+B11}$ | (Cibula et al., 1992) |
| Normalised Difference Snow Index | $NDSI = \frac{B3-B11}{B3+B11}$ | (Valovcin, 1976) |
| Normalised Difference Water Index | $NDWI = \frac{B3-B8}{B3+B8}$ | (Gao, 1996) |
| Bare Soil Index | $BSI = \frac{(B11+B4)-(B8+B2)}{(B11+B4)+(B8+B2)}$ | (Rikimaru et al., 2002) |
| Normalised Difference Salinity Index | $NDSI2 = \frac{B11-B12}{B11+B12}$ | (Richardson et al., 2002) |
| Burned Area Index for Sentinel-2 | $BAIS2 = (1 - \sqrt{\frac{B6*B7*B8A}{B4}}) * (\frac{B12-B8A}{\sqrt{B12+B8A}} + 1)$ | (Filipponi, 2018) |
| Inverted Red-Edge Chlorophyll Index | $IRECI = \frac{B7-B4}{\frac{B5}{B6}}$ | (Frampton et al., 2013) |

B.4 McNemar test p-values

Table B.3 – Classification comparison with the p-values derived from the McNemar significance test of performance on the test set. Values less than the significance level of $\alpha = 0.05$ indicate a statistically significant difference between the two models at the 95% confidence interval. Otherwise, if the p-values $> \alpha$ (highlighted in bold) the null hypothesis cannot be rejected.

| | Band 11 | Band 12 | PI ₂ | SIWSI | Bands | Indices | PI ₁ | PI ₃ | PI ₄ |
|-----------------|---------|---------|-----------------|--------------|---------|--------------|-----------------|-----------------|-----------------|
| Band 5 | < 0.001 | < 0.001 | < 0.001 | < 0.001 | 0 | 0 | < 0.001 | < 0.001 | < 0.001 |
| Band 11 | - | < 0.001 | < 0.001 | < 0.001 | < 0.001 | < 0.001 | < 0.001 | < 0.001 | < 0.001 |
| Band 12 | - | - | < 0.001 | 0.097 | < 0.001 | < 0.001 | < 0.001 | < 0.001 | 0.041 |
| PI ₂ | - | - | - | < 0.001 | < 0.001 | < 0.001 | < 0.001 | < 0.001 | < 0.001 |
| SIWSI | - | - | - | - | < 0.001 | < 0.001 | < 0.001 | < 0.001 | 0.014 |
| Bands | - | - | - | - | - | 0.395 | < 0.001 | < 0.001 | < 0.001 |
| Indices | - | - | - | - | - | - | < 0.001 | < 0.001 | < 0.001 |
| PI ₁ | - | - | - | - | - | - | - | < 0.001 | < 0.001 |
| PI ₃ | - | - | - | - | - | - | - | - | < 0.001 |

(a) 2017

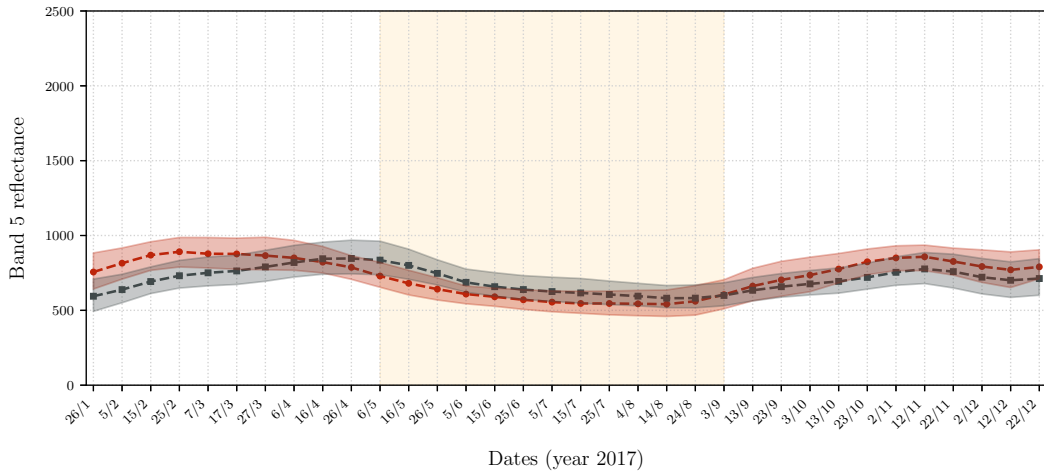
| | Band 11 | Band 12 | PI ₂ | SIWSI | Bands | Indices | PI ₁ | PI ₃ | PI ₄ |
|-----------------|---------|---------|-----------------|---------|--------------|---------|-----------------|-----------------|-----------------|
| Band 5 | < 0.001 | < 0.001 | < 0.001 | < 0.001 | < 0.001 | < 0.001 | < 0.001 | < 0.001 | < 0.001 |
| Band 11 | - | < 0.001 | < 0.001 | < 0.001 | < 0.001 | < 0.001 | < 0.001 | < 0.001 | 0.015 |
| Band 12 | - | - | < 0.001 | < 0.001 | < 0.001 | < 0.001 | 0.069 | < 0.001 | < 0.001 |
| PI ₂ | - | - | - | < 0.001 | 0.523 | < 0.001 | < 0.001 | 0 | < 0.001 |
| SIWSI | - | - | - | - | < 0.001 | < 0.001 | < 0.001 | < 0.001 | < 0.001 |
| Bands | - | - | - | - | - | < 0.001 | < 0.001 | 0 | < 0.001 |
| Indices | - | - | - | - | - | - | < 0.001 | 0 | < 0.001 |
| PI ₁ | - | - | - | - | - | - | - | < 0.001 | < 0.001 |
| PI ₃ | - | - | - | - | - | - | - | - | < 0.001 |

(b) 2018

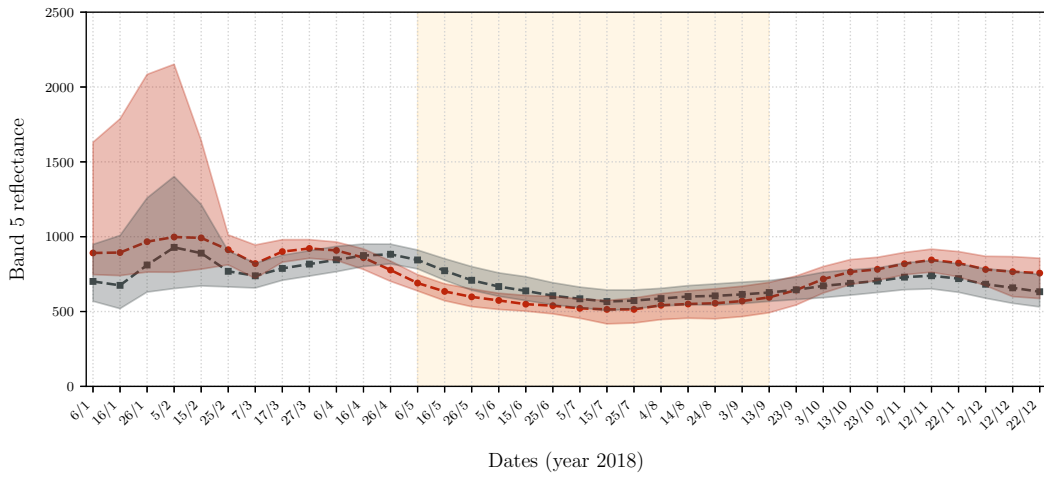
Table B.4 – Poplar PA scores achieved with the ten best feature subsets based on SFFS in the single-feature configuration using the three variants of the poplar index: PI_1 , PI_3 and PI_4 . The test score gives the poplar PA after prediction on the test set (50% of reference data). These values are significantly lower than the best PA scores obtained with PI_2 : 92% in 2017 and 95% in 2018 (see Table 6.3 in the core of the manuscript).

| Year | Poplar PA scores | PI_1 | PI_3 | PI_4 |
|------|------------------|--------|--------|--------|
| 2017 | test score | 84.9 | 73.9 | 88.3 |
| 2018 | test score | 89.7 | 73.3 | 92.2 |

B.5 Spectral signatures



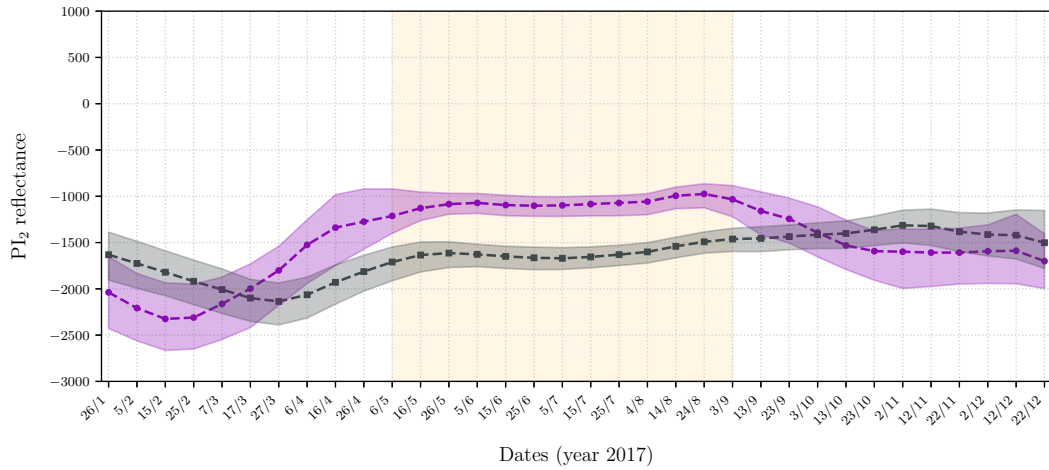
(a)



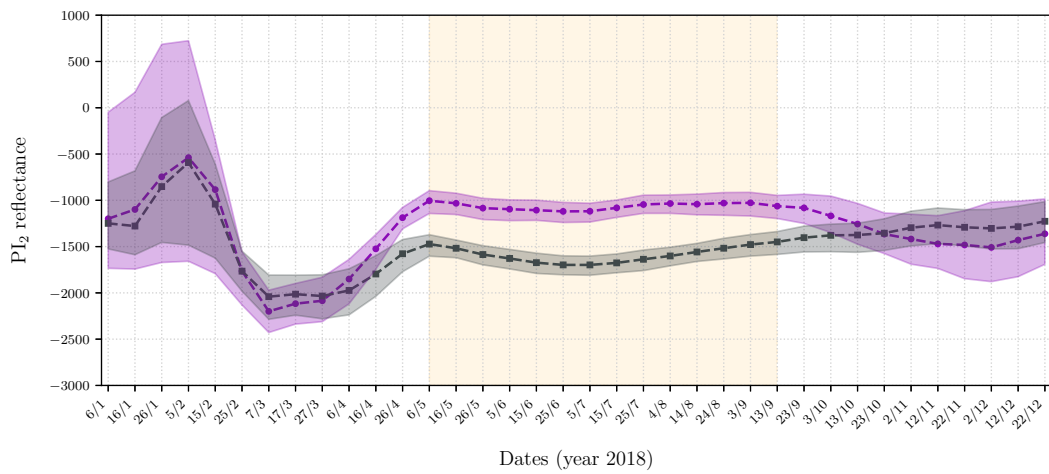
(b)



Figure B.2 – Temporal profile of band 5 reflectance in poplar plantations versus other deciduous tree species. In (a) the profile in 2017 and in (b) the profile in 2018. Dashed lines refer to the median signal and the surrounding shaded areas represent the interquartile range, shown in red for poplar and in grey for other deciduous species. The light yellow box highlights the most discriminating spring-summer period for poplars.



(a)



(b)



Figure B.3 – Temporal profile of PI_2 reflectance in poplar plantations versus other deciduous tree species. In (a) the profile in 2017 and in (b) the profile in 2018. Dashed lines refer to the median signal and the surrounding shaded areas represent the interquartile range, shown in purple for poplar and in grey for other deciduous species. The light yellow box highlights the most discriminating spring-summer period for poplars.

Annexe C

Compléments chapitre 7

C.1 Acquisitions Sentinel-2 utilisées en 2018

TABLEAU C.1 – Nombre d’acquisitions Sentinel-2 disponibles par tuile en 2018. Ce nombre reflète les emprises selon toutes les orbites relatives traversant chaque tuile.

| Tuile | Nombre d’acquisitions |
|--------------|------------------------------|
| 30TVT | 92 |
| 31UER | 100 |
| 31TFL | 95 |
| 31TEL | 80 |
| 30UXV | 84 |
| 32TLT | 89 |
| 30TXP | 74 |
| 30UWA | 92 |
| 31TFH | 120 |
| 30TYS | 73 |
| 31TCN | 74 |
| 31UGQ | 68 |
| 31UDQ | 71 |
| 30TXT | 64 |
| 30UYV | 102 |
| 31TDJ | 79 |
| 31TCJ | 94 |
| 30UXA | 91 |
| 32ULV | 71 |
| 31TFM | 90 |
| 32TLS | 146 |
| 32TLP | 105 |
| 31UDP | 71 |
| 30TXN | 101 |
| 31UDS | 82 |
| 30TYQ | 75 |
| 31UFP | 99 |
| 30UYA | 81 |
| 32TNM | 108 |

| Tuile | Nombre d'acquisitions |
|--------------|------------------------------|
| 31TDK | 81 |
| 32TMN | 152 |
| 32TLR | 116 |
| 30UYU | 92 |
| 31TFN | 103 |
| 31UFR | 69 |
| 31TFJ | 115 |
| 31UCQ | 97 |
| 32TNL | 113 |
| 31TDL | 81 |
| 30TYP | 75 |
| 32TMM | 107 |
| 32TLQ | 107 |
| 31UEQ | 78 |
| 31TEH | 114 |
| 32UMU | 80 |
| 30UVV | 84 |
| 31TDH | 98 |
| 30TXQ | 76 |
| 31TGK | 108 |
| 30TYT | 71 |
| 31TFK | 104 |
| 32TNN | 109 |
| 30UUU | 119 |
| 31UES | 109 |
| 30UWV | 89 |
| 31TEM | 82 |
| 31UGP | 75 |
| 31TDM | 116 |
| 30TYN | 101 |
| 31TGL | 112 |
| 30UVU | 70 |
| 30TYR | 72 |

| Tuile | Nombre d'acquisitions |
|--------------|------------------------------|
| 31UDR | 70 |
| 31TCK | 67 |
| 31TGH | 115 |
| 30TWP | 92 |
| 30TXS | 87 |
| 31TEN | 75 |
| 31TDN | 104 |
| 30UWU | 99 |
| 32UMV | 81 |
| 31TEJ | 143 |
| 30TWT | 96 |
| 31TGM | 101 |
| 31UFQ | 72 |
| 31TCL | 64 |
| 30UXU | 71 |
| 31UCP | 68 |
| 31TCH | 102 |
| 31UFS | 72 |
| 31UCS | 96 |
| 31TEK | 82 |
| 32ULU | 76 |
| 30TXR | 82 |
| 31UEP | 75 |
| 31TGN | 90 |
| 30TWS | 90 |
| 31UCR | 115 |
| 31TCM | 69 |
| 31TGJ | 151 |