Rapport Synthèse

Cartographie automatisée d'essences pour la région du Lac-St-Jean en vue d'une meilleure planification pour les prescriptions de traitement

Présenté à

Société de protection des forêts contre les insectes et maladies

Éric Litalien, ing.f.

Maria-Luisa Souto-Neveu

Par :

CERFO
Centre d'enseignement et de recherche en foresterie de Sainte-Foy inc.

Batistin Bour, M.Sc.

Mathieu Varin, M.Sc.

Martin Dupuis, M.Sc.

Mars 2021

Mots-clés : cartographie, Sentinel-2, lidar, apprentissage automatique, intelligence artificielle, télédétection, classification par pixel

Référence à citer :

Bour, B., Varin, M., Dupuis, M. 2021. Cartographie automatisée d'essences pour la région du Lac St-Jean en vue d'une meilleure planification pour les prescriptions de traitement. Centre d'enseignement et de recherche en foresterie de Sainte-Foy inc. (CERFO). Rapport 2021-15. 44 pages.

CERFO
Centre d'enseignement et de recherche en foresterie de Sainte-Foy inc.

Table des matières

1. Introduction....4
2. Territoire d'étude....5
3. Données utilisées....5
3.1. Images aériennes....5
3.2. Images satellitaire Sentinel-2....6
3.3. Données lidar....8
4. Méthode....9
4.1. Photo-interprétation....9
4.2. Préparation des images Sentinel-2....9
4.3. Variables générées....10
4.3.1. Variables spectrales....10
4.3.2. Variables lidar....12
4.4. Modélisation par apprentissage automatique....14
4.5. Application du modèle sur le territoire d'intérêt....15
5. Résultats....15
5.1. Statistiques exploratoires....15
5.2. Modèles retenus....17
5.2.1. Modèle cinq dates....19
5.2.2. Modèle deux dates....21
5.3. Application des modèles....23
5.3.1. Modèle cinq dates....23
5.3.2. Modèle deux dates....25
5.3.3. Comparaisons....27
5.3.4. Classification de l'épinette de Norvège....28
6. Discussion....29
6.1. Limites....29
6.2. Pistes d'amélioration....30
7. Références....32
8. Annexe : Guide d'utilisation....35

Liste des figures

Figure 1 . Situation géographique du territoire d'étude et sa composition en domaines bioclimatiques. 5
Figure 2 . Emplacement et référence des images Sentinel-2 nécessaires à recouvrir tout le territoire d'étude. 7
Figure 3. Présences et absence de feuilles lors des acquisitions lidar du territoire d'étude. 9
Figure 4 . Processus de traitement des images Sentinel-2. 10
Figure 5 . Nombre d'observations issues de la photo-interprétation (AU = autre résineux, EB = épinette blanche, EN = épinette noire, FE = Feuillus, SB = sapin baumier). 16
Figure 6 . Nombre d'observations issues de la méthode d'apprentissage actif (AU = autre résineux, EB = épinette blanche, EN = épinette noire, FE = Feuillus, SB = sapin baumier). 16
Figure 7 . Évolution temporelle de l'indice spectral NDVI selon les essences forestières d'intérêt (AU = autre résineux, EB = épinette blanche, EN = épinette noire, FE = Feuillus, SB = sapin baumier). 17
Figure 8 . Évolution des performances (Kappa de Cohen) selon le nombre de date intégrées dans le modèle random forest (JAUNE = Pas de sélection de variable; BLEU = Sélection de variables). 19
Figure 9 . Importance des variables basé sur le mean decrease accurary du modèle random forest développé à partir de cinq dates différentes. 21
Figure 10. Importance des variables basé sur le mean decrease accurary du modèle random forest développé à partir de 2 dates différentes. 23
Figure 13. Cartographie des essences forestière d'intérêt pour toutes les surfaces forestières du territoire d'étude avec le modèle utilisant 5 dates (AU = autre résineux, EB = épinette blanche, EN = épinette noire, FE = Feuillus, SB = sapin baumier). 24
Figure 14 . Cartographie des essences forestière d'intérêt pour toutes les surfaces forestières du territoire d'étude avec le modèle utilisant 2 dates (AU = autre résineux, EB = épinette blanche, EN = épinette noire, FE = Feuillus, SB = sapin baumier). 26
Figure 15. Comparaison entre les résultats de classification des modèles à cinq dates (a), deux dates (b) et la carte écoforestière (c), sur une portion du territoire d'intérêt. 27
Figure 16. Exemple de classifications des modèles deux dates (a) et cinq dates (b) en comparaison avec plusieurs points photo-interprétés. 28
Figure 17 . Classification des 12 épinettes de Norvège identifiées par photo-interprétation (AU = autre résineux, EB = épinette blanche, EN = épinette noire, FE = Feuillus, SB = sapin baumier). 29

Liste des tableaux

Tableau 1. Caractéristiques des bandes spectrales des images Sentinel-2 disponibles. 6
Tableau 2. Liste des images Sentinel-2 téléchargées et leur date d'acquisition.....7
Tableau 3. Caractéristiques des données lidar disponibles dans le territoire d'intérêt.....8
Tableau 4 . Liste des indices spectraux calculés.....11
Tableau 5 . Liste des métriques lidar calculées pour chaque nuage de points.....12
Tableau 6 . Performances enregistrées des modèles random forest avec plusieurs combinaisons de dates et de sélection de variables. 18
Tableau 7 . Matrice de confusion du modèle random forest développé avec 5 dates présentant le meilleur score de Kappa (AU = autre résineux, EB = épinette blanche, EN = épinette noire, FE = Feuillus, SB = sapin baumier). 20
Tableau 8 . Matrice de confusion du modèle random forest développé avec 2 dates présentant le meilleur score de Kappa (AU = autre résineux, EB = épinette blanche, EN = épinette noire, FE = Feuillus, SB = sapin baumier).....21

1. Introduction

Les forêts du Québec sont soumises à une grande variété de perturbations naturelles (Gagnon et Morin, 2001). Parmi celles-ci, les ravageurs forestiers sont une source majeure de perturbations. La tordeuse du bourgeon d'épinette (TBE : *Choristoneura fumiferana*) est l'insecte causant le plus de dommages aux forêts (Bouchard, 2011). Au Québec, la TBE s'attaque principalement au sapin baumier (*Abies balsamea*) et à l'épinette blanche (*Picea glauca*), mais peut également engendrer des dommages chez l'épinette noire (*Picea mariana*) (Bouchard, 2011). Les populations de TBE suivent un cycle épidémique étendu sur 30 ans (Girard *et al.*, 2008). Les pics épidémiques ont été documentés au Québec depuis le début du 20ème siècle et ont touché à ce jour près de 88 Mha de surfaces forestières (Bouchard, 2011). Afin de limiter la perte de ressource ligneuse, le Québec mène une gestion intégrée des épidémies de TBE par l'aménagement préventif des forêts et par lutte directe (Bouchard, 2011). La société de protection des forêts contre les insectes et maladies (SOPFIM) est chargée de la planifier et met en œuvre des moyens de lutte contre la TBE. L'arrosage des peuplements forestiers avec l'insecticide biologique *Bacillus thunringiensis* var. *kurstaki* (BTK; Fuentealba *et al.*, (2015)) est le seul moyen de lutte directe utilisé au Québec. Le processus de sélection des aires admissibles au traitement par BTK prend en compte l'abondance de TBE, les contraintes aux opérations forestières et la vulnérabilité des peuplements forestiers. La vulnérabilité est entre autres définie par le type d'essence forestière présente.

Bien que des données écoforestières issues de la photo-interprétation soient disponibles sur tout le Québec (MFFP, 2018), l'échelle d'analyse (unité minimale cartographique : 4 ha) ne permet pas d'identifier avec précision les peuplements les plus intéressants pour l'utilisation du BTK. Ce dernier reste onéreux et requiert des visites terrain pour confirmer la présence d'essences forestières d'intérêt préalablement identifiées dans les données écoforestières (Fuentealba *et al.*, 2015).

Les données géospatiales de toutes sortes (plateformes, résolutions spatiale, temporelle et radiométrique) sont de plus en plus rependues et accessibles. L'imagerie satellitaire constitue une source de données périodique sur de vastes étendues (Boucher *et al.*, 2015; Shojanoori et Shafri, 2016; Wulder *et al.*, 2016). Depuis l'année 2017, l'imagerie Sentinel-2, une série de satellites d'observation de la Terre de l'Agence spatiale européenne (ESA), fournie de l'information spectrale tous les 5 jours (Grabska *et al.*, 2019). Plusieurs études ont montré la capacité des images Sentinel-2 à classifier des essences forestières (Grabska *et al.*, 2019; Immitzer *et al.*, 2019). L'utilisation de l'imagerie Sentinel-2 pour identifier les essences forestières les plus susceptibles d'être touchées par la TBE permettrait un gain de temps et de ressources considérables pour identifier les aires admissibles au traitement par BTK. De plus, la résolution spatiale de 10 m des images Sentinel-2 est plus fine que les polygones de la carte écoforestière. Possédant une très fine résolution spatiale, les données lidar permettent d'acquérir des informations tridimensionnelles sur les peuplements forestiers, pouvant servir au calcul d'un grand nombre de métriques caractérisant leur structure. Tout comme l'imagerie, les données lidar ont été utilisées à plusieurs reprises dans des projets visant la classification d'essences forestières (Alonzo *et al.*, 2014; White *et al.*, 2013). Une acquisition

de données lidar de tout le Québec méridional est en cours d'acquisition et devrait être complétée en 2022 (Leboeuf *et al.*, 2015). L'utilisation de données géospatiales par télédétection présentent donc un grand potentiel pour identifier des essences d'intérêts, pouvant aider les gestionnaires des forêts du Québec dans le processus de sélection des aires admissibles à la lutte contre la TBE.

2. Territoire d'étude

Le territoire d'étude est situé dans la région du Saguenay-Lac-St-Jean (Figure 1). La superficie totale est d'approximativement 500 000 ha. Le territoire est principalement composé de surfaces forestières et de zones agricoles présentes aux abords du Lac-Saint-Jean. Une proportion importante de terres privées est retrouvée dans le territoire d'étude. La disponibilité de données lidar sert de contrainte délimitant les bordures du territoire d'étude.

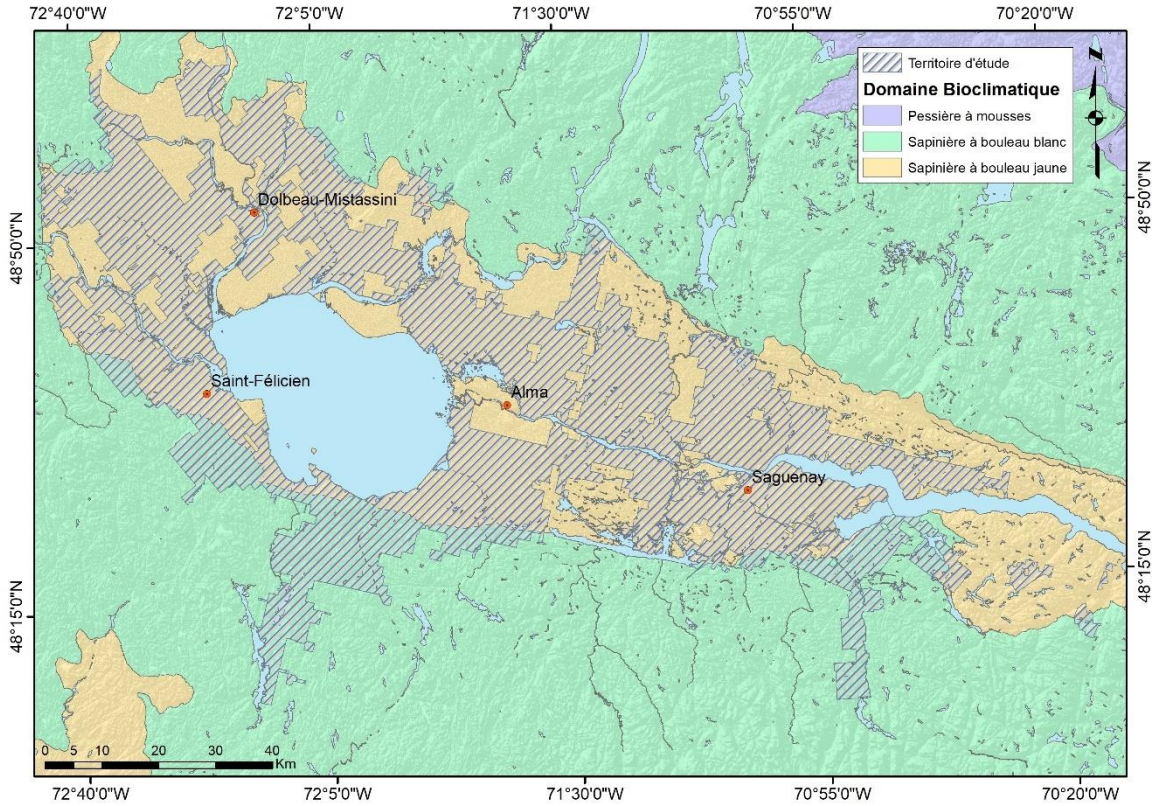


Figure 1 . Situation géographique du territoire d'étude et sa composition en domaines bioclimatiques.

3. Données utilisées

3.1. Images aériennes

Des images aériennes couvrant tout le territoire d'étude ont été obtenues auprès de la direction des inventaires forestiers. Les images aériennes sont acquises dans le cadre de l'inventaire décennal des forêts du Québec. Les photos aériennes couvrant la portion est du

territoire ont été acquises au printemps avant l'apparition des feuilles, tandis que les photos de la portion ouest ont été acquises pendant la saison de croissance.

3.2. Images satellitaire Sentinel-2

Les images Sentinel-2 utilisées présentent 10 bandes spectrales (Tableau 1) allant du bleu jusqu'à l'infrarouge à onde courte.

Tableau 1. Caractéristiques des bandes spectrales des images Sentinel-2 disponibles.

Bande	Spectre	Largeur de la bande (nm)	Centre de la bande (nm)	Résolution spatiale (m)
2	Bleu	66	492.4	10
3	Vert	36	559.8	10
4	Rouge	31	664.6	10
5	Rouge	15	704.1	20
6	Red-Edge	15	740.5	20
7	Proche-infrarouge 1	20	782.8	20
8	Proche-infrarouge 2	106	832.8	10
8a	Proche-infrarouge 3	21	864.7	20
11	Infrarouge onde courte 2	91	1613.7	20
12	Infrarouge onde courte 3	175	2202.4	20

Afin de couvrir l'ensemble du territoire d'étude, plusieurs tuiles Sentinel-2 ont été nécessaires (Figure 2). Plusieurs dates ont été acquises (mai, juin, juillet, septembre et octobre). Afin d'avoir une couverture sans nuages de tout le territoire d'étude, il a été nécessaire de récupérer chaque mois désiré avec des années différentes (Tableau 2). De plus, pour le mois de septembre et octobre, plusieurs années ont été nécessaires pour couvrir l'ensemble du territoire d'étude sans avoir de nuages.

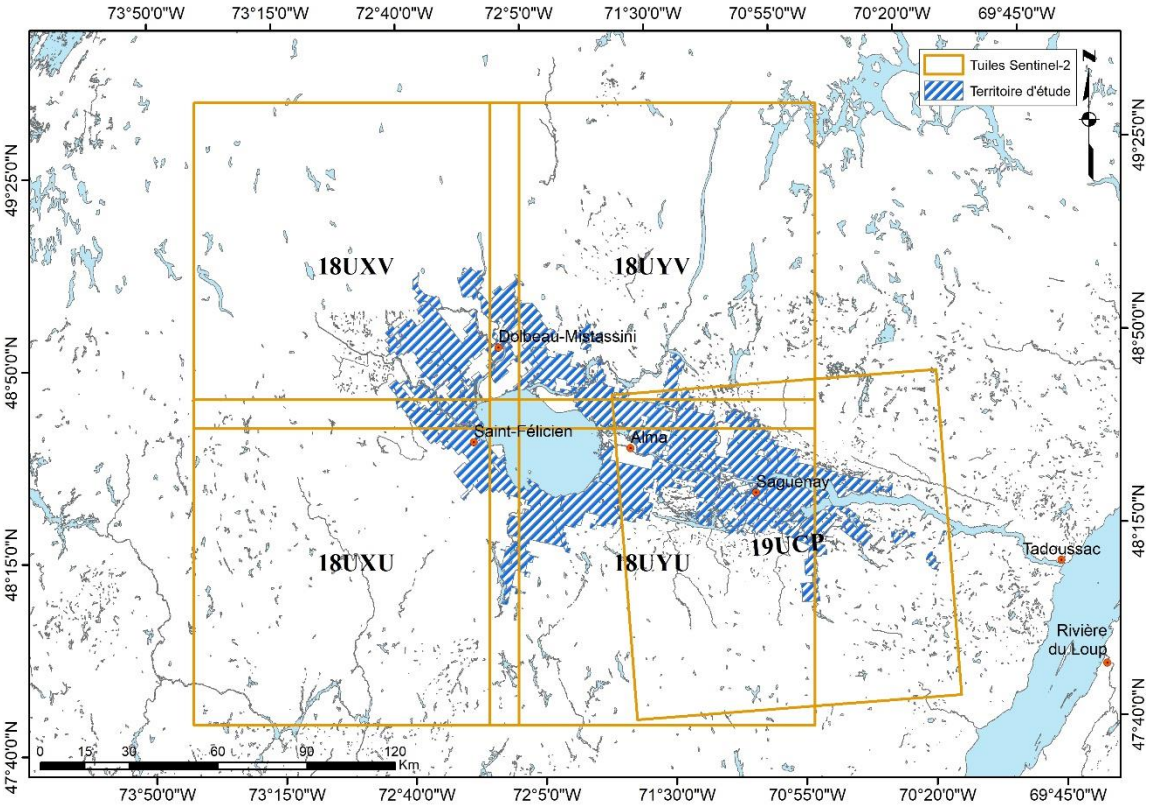


Figure 2 . Emplacement et référence des images Sentinel-2 nécessaires à recouvrir tout le territoire d'étude.

Tableau 2. Liste des images Sentinel-2 téléchargées et leur date d'acquisition.

Mois	Numéro de tuile	Date d'acquisition
Mai	T19UCP	18-05-2018
	T18UXU	26-05-2018
	T18UXV	26-05-2018
	T18UYU	26-05-2018
	T18UYV	26-05-2018
	T19UCP	26-05-2018
Juin	T18UXU	09-06-2020
	T18UXV	09-06-2020
	T18UYU	09-06-2020
	T18UYV	09-06-2020
	T19UCP	16-06-2020
	T18UYV	21-06-2020
Juillet	T18UXU	15-07-2019
	T18UXV	15-07-2019
	T18UYV	10-07-2019
	T19UCP	10-07-2019
	T18UYU	27-07-2019
	T19UCP	27-07-2019

Septembre	T19UCP	20-09-2017
	T18UXU	13-09-2018
	T18UXV	13-09-2018
	T18UYU	13-09-2018
	T18UYV	13-09-2018
	T19UCP	13-09-2018
Octobre	T18UXU	23-10-2017
	T19UCP	10-10-2019
	T19UCP	10-10-2019
	T18UXV	12-10-2020
	T18UYU	12-10-2020
	T18UYV	12-10-2020
	T19UCP	12-10-2020

3.3. Données lidar

Plusieurs acquisitions lidar, à des années différentes, ont été nécessaires pour couvrir l'ensemble du territoire d'étude (Tableau 3). La densité de points varie entre 2 et 4 points par mètre carré. Une majorité des acquisitions lidar ont été réalisées lors d'absence de feuilles dans les arbres (Figure 3).

Tableau 3. Caractéristiques des données lidar disponibles dans le territoire d'intérêt.

Propriétaire	Année	Densité (pts/m^2)	Présence de feuilles
Ministère des Transports du Québec	2013	2	NON
Ministère des Transports du Québec	2017	4	NON
Ministère des Transports du Québec	2018	4	NON
Ministère des Transports du Québec	2013	2	NON
Ministère des Transports du Québec	2016	4	NON
Ministère des Transports du Québec	2016	4	NON
Direction des Inventaires Forestiers	2019	2.5	OUI
Direction des Inventaires Forestiers	2019	2.5	OUI
Soutien aux Opérations Régionales	2017	2.5	OUI
Direction des Inventaires Forestiers	2016	2.5	OUI
Soutien aux Opérations Régionales	2017	2.5	OUI

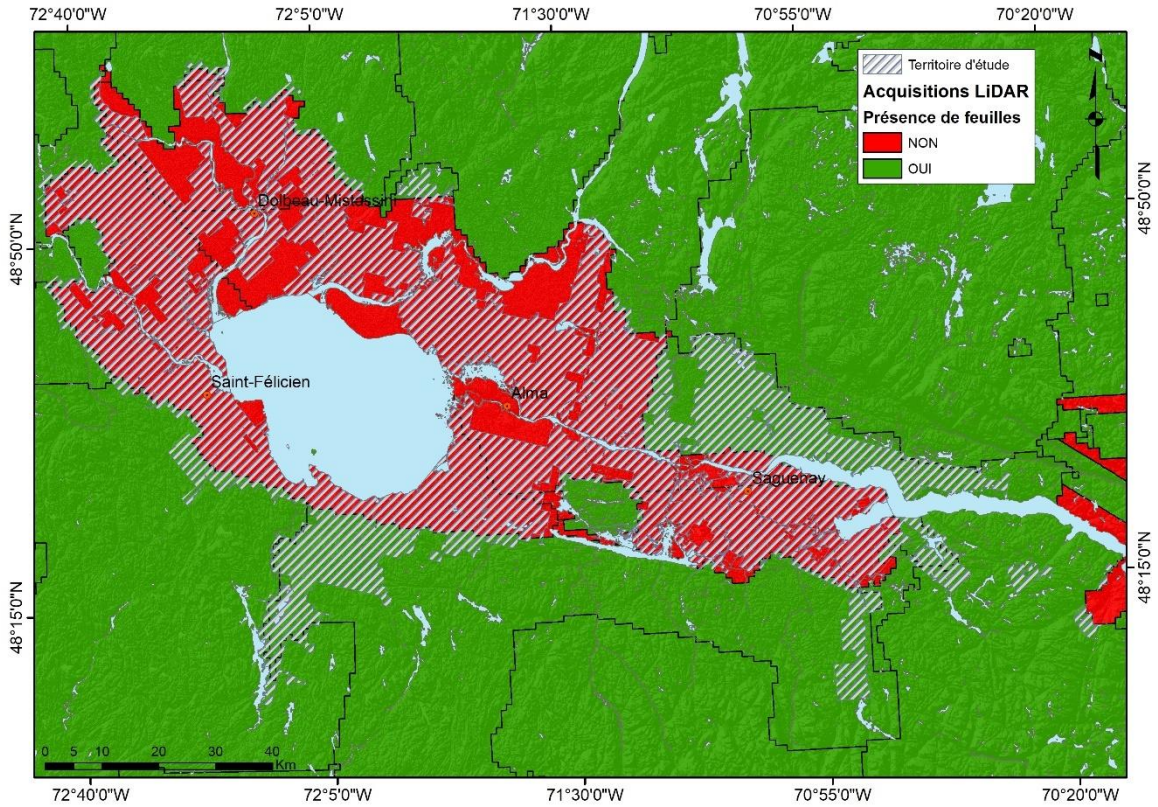


Figure 3. Présences et absence de feuilles lors des acquisitions lidar du territoire d'étude.

4. Méthode

4.1. Photo-interprétation

Les images aériennes ont été utilisées pour identifier les essences d'intérêt. Une grille spatiale correspondant aux pixels des images Sentinel-2 a été utilisée pour identifier des pixels d'essence pure. Initialement 7 classes étaient recherchées soit : Épinette noire, Épinette blanche, Épinette de Norvège, Sapin baumier, Autres résineux, Feuillus et Arbres morts. Un total de 428 observations a ainsi été identifié. Très peu d'observations d'épinette de Norvège ont été relevées puisque peu d'information était disponible en terres privées, dont la proportion est importante dans le territoire d'étude. La présence d'arbres morts était trop faible pour remplir entièrement des pixels des images Sentinel-2. De plus, les images acquises au printemps rendaient difficile la distinction entre arbres morts et arbres feuillus sans feuilles.

4.2. Préparation des images Sentinel-2

Les images Sentinel-2 ont été téléchargées à partir du portail EarthExplorer (https://earthexplorer.usgs.gov/). Plusieurs produits sont fournis par l'ESA, cependant les images avec traitements atmosphériques (niveau L2A) sont archivées un an après leur acquisition. Pour avoir accès à ces données, il est nécessaire de faire plusieurs demandes avec un quota de téléchargement rapidement atteint. Pour contourner cette contrainte, les images

ont été téléchargées en format sans corrections atmosphériques (niveau L1C) et leur traitement a été effectué dans le logiciel SNAP créé par l'ESA. Les images Sentinel-2 L2A ont ensuite été importées dans le logiciel R (R Core Team, 2020) pour subir une série de traitements. D'abord, les zones avec nuages ont été retirées des images. Ensuite, le territoire d'étude a été utilisé pour masquer les images, ce qui réduit considérablement le temps de traitement pour les étapes subséquentes grâce à la diminution du nombre de pixels à traiter. L'étape suivante a pour but d'obtenir toutes les images dans une même projection. Les images reprojetées ont subi un rééchantillonnage bilinéaire. Enfin, la dernière étape consiste à créer une mosaïque de l'ensemble du territoire d'étude avec toutes des images. Une mosaïque est produite pour chaque mois et pour chaque bande spectrale. Les différents traitements ont été synthétisés dans la Figure 4.

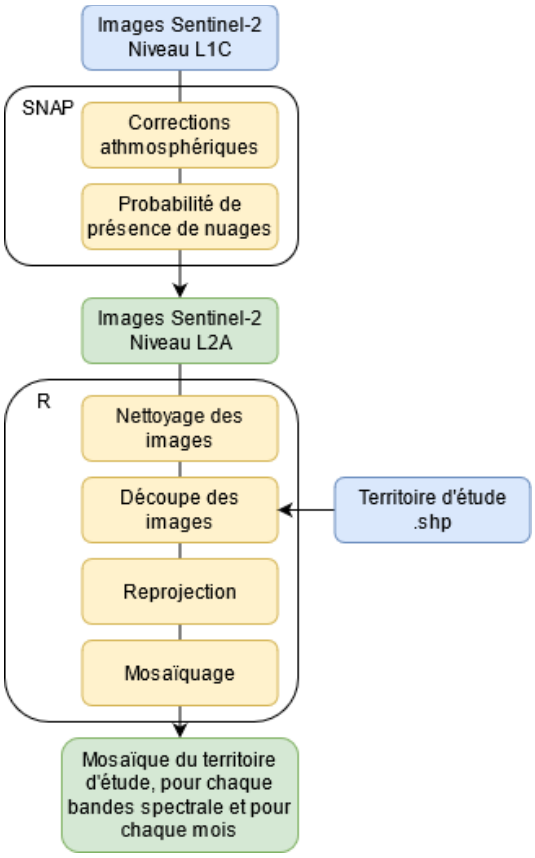


Figure 4 . Processus de traitement des images Sentinel-2.

4.3. Variables générées

4.3.1. Variables spectrales

Pour chaque observation identifiée lors de la photo-interprétation, l'ensemble des valeurs spectrales, pour chaque mois et chaque bandes spectrales ont été extraites dans le logiciel R. Les bandes spectrales ont été utilisées pour calculer 70 indices spectraux listés dans le Tableau 4.

Tableau 4 . Liste des indices spectraux calculés.

Indice spectral	Équation	Référence
ABDI1_B08a_B11	B08a ÷ atan((B08a - B11+0.001)÷(1613.7-864.7))	(Liu *et al.*, 2010)
ABDI1_B08a_B12	B08a ÷ atan((B08a - B12)÷(2202.4-864.7))	(Liu *et al.*, 2010)
AWEI	(4*(B03-B11)-(0.25*B08+2.75*B12))÷(B03+B08+B11+B12)	(Mustafa *et al.*, 2017)
BAI	(B02 - B08)÷(B02 + B08)	(Chuvieco *et al.*, 2002)
CARI	B05÷B04*abs((B05-B03)÷150*670+B04+(B03-((B05-B03)÷150*550)))÷sqrt((B05-B03÷150^2+1))	(Xue et Su, 2017)
CARI2	abs((B05-B03)÷150*4+4)÷sqrt(2)*5÷4	(Xue et Su, 2017)
CG	B03÷B07	(Kobayashi *et al.*, 2020)
CRE	B05÷B07	(Kobayashi *et al.*, 2020)
CTVI	((B05-B03)÷(B05+B03)+0.5)÷abs(((B05-B03)÷(B05+B03)+0.5))*sqrt(abs(((B05-B03)÷(B05+B03)+0.5)))	(Morcillo-Pallarés *et al.*, 2019)
Datt1	(B08-B05)÷(B08-B04)	(Marrs et Ni-Meister, 2019)
Datt6	B04÷(B03*B05)	(Marrs et Ni-Meister, 2019)
Diff_B08_B03	B08a÷(B03*B05)	(Kobayashi *et al.*, 2020)
Diff_B08_B04	B08-B03	(Kobayashi *et al.*, 2020)
DVI	B08-B04	(Xue et Su, 2017)
GI	B03-B04	(Kobayashi *et al.*, 2020)
GNDVI	(B08-B03)÷(B08+B03)	(Xue et Su, 2017)
LCI	(B08-B05)÷(B08+B04)	(Kobayashi *et al.*, 2020)
LCCI	B07÷B06	(Kobayashi *et al.*, 2020)
Maccioni	(B07-B05)÷(B07-B04)	(Kobayashi *et al.*, 2020)
MCARI	((B05-B04)-0.2*(B05-B03))*B05÷B04	(Xue et Su, 2017)
MCARI1	1.2*(2.5*(B08-B04)-1.3*(B08-B03))	(Xue et Su, 2017)
MCARI2	1.5*(2.5*(B08-B04)-1.3*(B08-B03))÷sqrt((2*B08+1)^2-(6*B08-5*sqrt(B04)-0.5))	(Xue et Su, 2017)
MSAVI	0.5*((2*B08+1)-sqrt((2*B08+1)^2-8*(B08-B04)))	(Xue et Su, 2017)
MSI	B11÷B08	(Xue et Su, 2017)
MTVI1	1.2*(1.2*(B08-B03)-2.5*(B04-B03))	(Xue et Su, 2017)
MTVI2	1.5*(1.2*(B08-B03)-2.5*(B04-B03))÷sqrt((2*B08+1)^2-(6*B08-5*sqrt(B04)-0.5))	(Xue et Su, 2017)
mNDWI_B03_B11	(B03-B11)÷(B11+B03)	(Kobayashi *et al.*, 2020)
ND_B7_B03	(B07-B03)÷(B07+B03)	(Xue et Su, 2017)
ND_B7_B04	(B07-B04)÷(B07+B04)	(Xue et Su, 2017)
ND_B08_B02	(B08-B02)÷(B08+B02)	(Xue et Su, 2017)
ND_B08_B03	(B08-B03)÷(B08+B03)	(Xue et Su, 2017)
ND_B08_B11	(B08-B11)÷(B08+B11)	(Xue et Su, 2017)
ND_B03_B05	(B03-B05)÷(B03+B05)	(Xue et Su, 2017)
ND_B12_B08	(B12-B08)÷(B12+B08)	(Xue et Su, 2017)
ND_B08_B04	(B08-B04)÷(B08+B04)	(Xue et Su, 2017)
ND_B05_B03	(B05-B03)÷(B05+B03)	(Xue et Su, 2017)
ND_B08a_B12	(B08a-B12)÷(B08a+B12)	(Xue et Su, 2017)
NDMI	(B08a-B11)÷(B08a+B11)	(Chance *et al.*, 2017)

NDRE_SWIR	(B06-B12)÷(B06+B12)	(Kobayashi *et al.*, 2020)
NDSI	(B11-B08)÷(B08+B11)	(Kobayashi *et al.*, 2020)
NDTI	(B11-B12)÷(B11+B12)	(Quemada et Daughtry, 2016)
NDVI	(B08-B04)÷(B08+B04)	(Gamon *et al.*, 1995)
NDWI	(B03-B08)÷(B08+B03)	(Mustafa *et al.*, 2017)
NDWI_B03_B08a	(B03-B08a)÷(B08a+B03)	(Mustafa *et al.*, 2017)
NHI	(B11-B02)÷(B11+B02)	(Pimstein *et al.*, 2009)
NMDI2	(B08a-B11-B12)÷(B08a+B11+B12)	(Xue et Su, 2017)
OSAVI	1.16*(B08-B04)÷(B08+B04+0.16)	(Xue et Su, 2017)
RDVI	(B08-B04)÷sqrt(B08+B04)	(Xue et Su, 2017)
REIP1	700+40*(((B04+B07)÷2-B05)÷(B06-B05))	(Kobayashi *et al.*, 2020)
REIP2	702+40*(((B04+B07)÷2-B05)÷(B06-B05))	(Kobayashi *et al.*, 2020)
REIP3	705+35*(((B04+B07)÷2-B05)÷(B06-B05))	(Kobayashi *et al.*, 2020)
RVI	B08÷B04	(Xue et Su, 2017)
SAVI	(1+0.5)*(B08-B04)÷(B08+B04+0.5)	(Xue et Su, 2017)
SR_BRE1	B02÷B05	(Xue et Su, 2017)
SR_BRE2	B02÷B06	(Xue et Su, 2017)
SR_BRE3	B02÷B07	(Xue et Su, 2017)
SR_NIRB	B08÷B02	(Xue et Su, 2017)
SR_NIRG	B08÷B03	(Xue et Su, 2017)
SR_NIRR	B08÷B04	(Xue et Su, 2017)
SR_NIRRE1	B08÷B05	(Xue et Su, 2017)
SR_NIRRE2	B08÷B06	(Xue et Su, 2017)
SR_NIRRE3	B08÷B07	(Xue et Su, 2017)
SR_B12_B05	B12÷B07	(Xue et Su, 2017)
STI	B11÷B12	(Ermolaev et al., 2020)
SWI	B02+B03-B08	(Dixit *et al.*, 2019)
TVI	sqrt((B05-B03)÷(B05+B03)+0.5)	(Xue et Su, 2017)
TCI	1.2*(B05-B03)-1.5*(B04-B03)*sqrt(B05÷B04)	(Kobayashi *et al.*, 2020)
VI700	(B05-B04)÷(B05+B04)	(Kobayashi *et al.*, 2020)
WBI	(B02-B04)÷(B02+B04)	(Xue et Su, 2017)
WRI	(B03+B04+B08)÷B02	(Mustafa *et al.*, 2017)

4.3.2. Variables lidar

Pour chaque pixel d'essences identifiées, le nuage de points correspondant a été extrait dans le logiciel R. Plusieurs métriques lidar ont été calculées pour caractériser le nuage de point, celles-ci sont listés dans le Tableau 5. Après analyse, l'intensité des retours lidar n'a pas été utilisée, car aucune normalisation n'est présente entre les différents vols réalisés.

Tableau 5 . Liste des métriques lidar calculées pour chaque nuage de points.

Métrique	Retours utilisés
Total de retours	Tous les retours
Total des premiers retours	Premiers retours
Total de retours au-dessus de 1,37 m	Tous les retours

Nombre de premiers retours au-dessus de 1,37 m	Premiers retours
Nombre de seconds retours au-dessus de 1,37 m	Second retour
Nombre de troisièmes retours au-dessus de 1,37 m	Troisièmes retours
Nombre de quatrièmes retours au-dessus de 1,37 m	Quatrièmes retours
Nombre de cinquièmes retours au-dessus de 1,37 m	Cinquièmes retours
Nombre de sixièmes retours au-dessus de 1,37 m	Sixièmes retours
Nombre de septièmes retours au-dessus de 1,37 m	Septièmes retours
Nombre de huitièmes retours au-dessus de 1,37 m	Huitièmes retours
Nombre de neuvièmes retours au-dessus de 1,37 m	Neuvièmes retours
Hauteur minimum	Tous les retours
Hauteur maximum	Tous les retours
Hauteur moyenne	Tous les retours
Hauteur mode	Tous les retours
Hauteur médiane	Tous les retours
Écart-type de la hauteur	Tous les retours
Variance de la hauteur	Tous les retours
Coefficient de variation de la hauteur	Tous les retours
Kurtosis de la hauteur	Tous les retours
Skewness de la hauteur	Tous les retours
1er percentile de hauteur	Tous les retours
5ème percentile de hauteur	Tous les retours
10ème percentile de hauteur	Tous les retours
15ème percentile de hauteur	Tous les retours
20ème percentile de hauteur	Tous les retours
25ème percentile de hauteur	Tous les retours
30ème percentile de hauteur	Tous les retours
35ème percentile de hauteur	Tous les retours
40ème percentile de hauteur	Tous les retours
45ème percentile de hauteur	Tous les retours
50ème percentile de hauteur	Tous les retours
55ème percentile de hauteur	Tous les retours
60ème percentile de hauteur	Tous les retours
65ème percentile de hauteur	Tous les retours
70ème percentile de hauteur	Tous les retours
75ème percentile de hauteur	Tous les retours
80ème percentile de hauteur	Tous les retours
85ème percentile de hauteur	Tous les retours
90ème percentile de hauteur	Tous les retours
95ème percentile de hauteur	Tous les retours
99ème percentile de hauteur	Tous les retours
Canopy Relief Ratio	Tous les retours
Pourcentage de retours au-dessus de 1.37 m	Tous les retours
Pourcentage de premiers retours au-dessus de 1.37m	Premiers retours
Nombre de retours au-dessus de 1.37m	Tous les retours
Pourcentage de premiers retours au-dessus du mode	Premiers retours
Pourcentage de premiers retours au-dessus de la moyenne	Premiers retours

Pourcentage de retours au-dessus du mode	Tous les retours
Pourcentage de retours au-dessus de la moyenne	Tous les retours
Nombre de premiers retours au-dessus de la moyenne	Premiers retours
Nombre de premiers retours au-dessus du mode	Premiers retours
Nombre de retours au-dessus de la moyenne	Tous les retours
Nombre de retours au-dessus du mode	Tous les retours

4.4. Modélisation par apprentissage automatique

La base de données comprenant les variables spectrales et les variables lidar a été séparée en données d'entrainement (80 %) et de validation (20 %) de manière aléatoire spatialement. Les données d'entrainement ont été utilisées pour développer un modèle d'apprentissage machine de type random Forest (Breiman, 2001), intégré dans R avec le package *randomForest* (Liaw et Wiener, 2002). Un processus de sélection de variables a été réalisé avec le package *VSURF* (Genuer *et al.*, 2019). Les variables sélectionnées ont été utilisées pour entraîner un modèle random Forest et les hyperparamètres ont été ajustés tel que suggéré par Probst et al. (2019). Le modèle développé a été appliqué sur des données de validation, permettant d'évaluer ses performances avec le coefficient Kappa de Cohen (Cohen, 1960).

Après plusieurs tests, il a été constaté que le lidar était très bruité pour identifier l'essence forestière dû aux multiples acquisitions avec et sans feuilles, ainsi qu'avec une variété de densités de points. Bien que l'étape de photo-interprétation avait pour but d'identifier le maximum d'observations possibles, le faible nombre d'observations pour les essences les moins communes a eu d'important effets sur les performances du modèle.

Pour pallier au manque d'observations, une méthode d'apprentissage actif a été réalisée uniquement avec les bandes spectrales Sentinel-2. La première version du modèle développé, bien que n'atteignant pas un seuil de performances acceptable (kappa > 0,7 – concordance jugée très bonne; Monserud et Leemans, (1992)), a été appliqué à l'ensemble du territoire d'étude. Ensuite, les pixels dont la classification correspondait avec les informations contenues dans les polygones d'essences pures de la carte écoforestière ont été conservés. Parmi ces pixels, un échantillonnage a été effectué pour s'assurer d'avoir au moins 250 m entre les observations pour éviter l'autocorrélation spatiale. Les pixels ainsi sélectionnés ont été ajoutés aux pixels issus de l'étape de photo-interprétation, augmentant ainsi le nombre d'observations de 428 à 7136. Ce nouveau jeu de données a été séparé en données d'entrainement (80 %) et de validation (20 %). Le processus de sélection de variables et d'entraînement de modèles a été appliqué sur les nouvelles données d'entrainement. Plusieurs combinaisons de dates ont été testées afin évaluer l'apport de performance amené par l'intégration de données à plusieurs étapes phénologiques des arbres pendant la saison de croissance et de sénescence. Finalement, les performances ont été évaluées sur les nouvelles données de validation. Cette évaluation a été effectuée avec la matrice de confusion, qui nous informe sur deux types d'erreurs. Dans le premier cas (erreur d'omission), il s'agit d'évaluer si des pixels appartenant à une classe de référence ont été correctement classifiés par la prédiction. Dans le second cas (erreur de commission), on y détermine si les pixels classifiés

par la prédiction appartiennent vraiment à ces classes. Avec cette matrice, la précision globale, qui mesure le pourcentage de pixels correctement classifiés (diagonale de la matrice) peut être calculé. Également, le coefficient Kappa de Cohen est une mesure de concordance tenant compte à la fois des erreurs de commission et d'omission afin de mesurer la force de la classification par rapport à une classification totalement due au hasard (Hagen-Zanker, 2006).

4.5. Application du modèle sur le territoire d'intérêt

Suivant l'analyse des performances des modèles, le meilleur modèle a été sélectionné ainsi qu'un second modèle plus simple d'utilisation. Les variables nécessaires à l'application des modèles sur l'ensemble du territoire d'étude ont été calculées dans le logiciel R à partir des mosaïques de bandes spectrales. Les modèles ont ainsi pu être appliqués mur-à-mur aux surfaces forestières identifiées avec la carte écoforestière (exclusion des zones anthropiques, agricoles et humides), fournissant des cartographies des essences forestières.

5. Résultats

5.1. Statistiques exploratoires

La photo-interprétation a permis d'identifier près de 70 observations pour l'épinette noire, le sapin baumier et les feuillus, et près de 50 observations pour l'épinette blanche et les autres résineux (Figure 5). Le territoire d'étude étant à la limite nord de l'aire de répartition de l'épinette blanche, il a été difficile de trouver des pixels purs de cette essence.

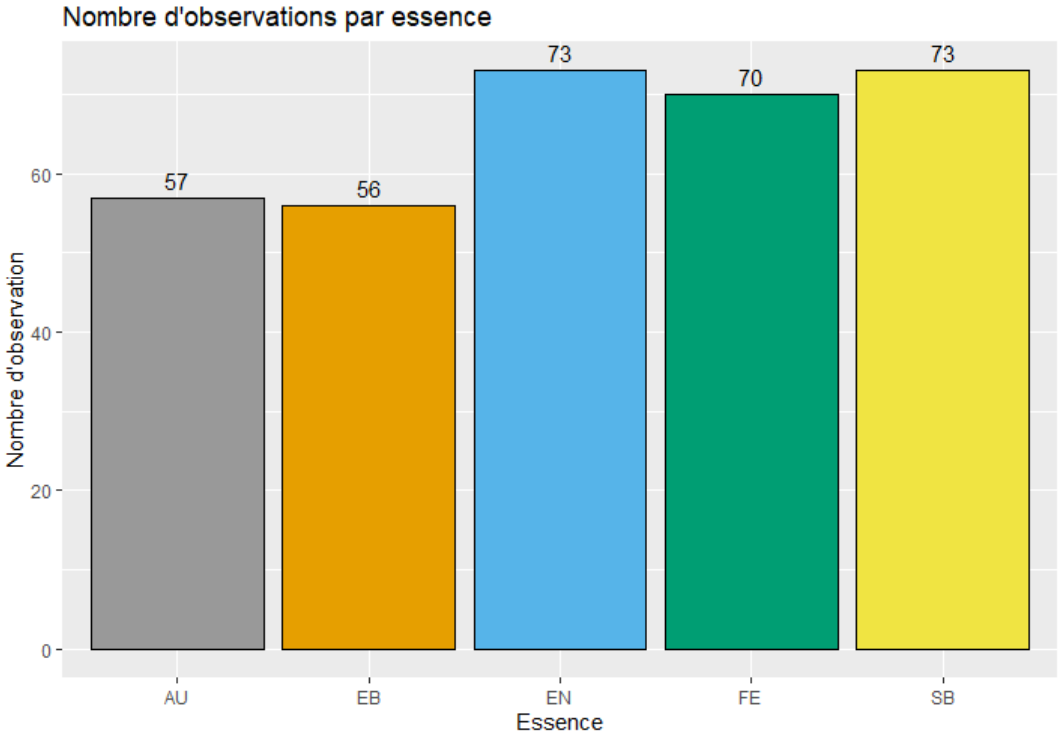


Figure 5 . Nombre d'observations issues de la photo-interprétation (AU = autre résineux, EB = épinette blanche, EN = épinette noire, FE = Feuillus, SB = sapin baumier).

L'application de la méthode d'apprentissage actif a permis d'augmenter considérablement le nombre d'observations à disposition (Figure 6). Cependant le nombre d'observation est débalancé avec la classe « Autre résineux » et « épinette noire » qui comptent beaucoup plus d'observations que les classes « épinette blanche », « feuillus » et « sapin baumier ». La faible occurrence de peuplements purs d'épinette blanche dans le territoire d'étude montre encore un effet sur le nombre d'observations disponibles. La faible proportion d'observations d'essences feuillues, pourtant largement présente dans le territoire d'étude, est attribuable aux catégories prises (essences individuelles pures) en compte dans la carte écoforestière. Cependant la signature spectrale des feuillus étant pressentie comme très différente des autres classes, il n'a pas été jugé nécessaire d'ajouter des observations supplémentaires pour balancer les classes. Le nombre d'observation de sapin baumier, bien que supérieur aux deux classes précédemment mentionnées, est bien en dessous de l'épinette noire et des autres résineux. Le sapin baumier étant une essence forestière de fin de succession écologique il est possible que l'identification de surface de 10 m x 10 m (taille du pixel) de sapin pur soit plus rare que les autres classes.

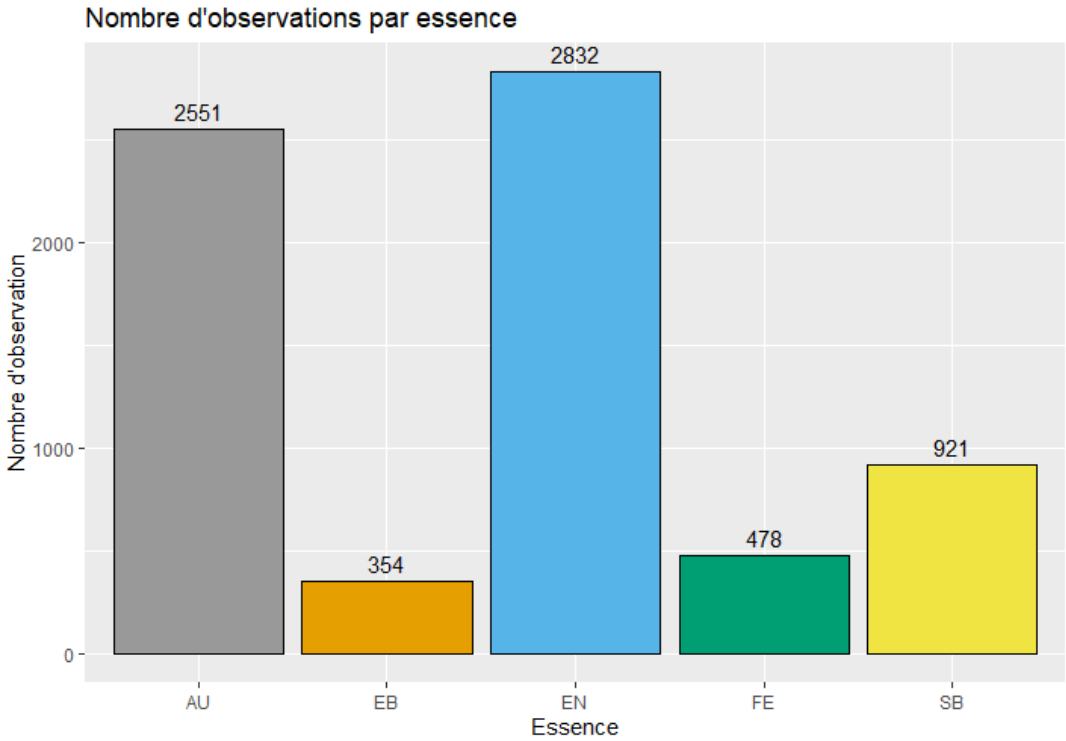


Figure 6 . Nombre d'observations issues de la méthode d'apprentissage actif (AU = autre résineux, EB = épinette blanche, EN = épinette noire, FE = Feuillus, SB = sapin baumier).

Les valeurs spectrales extraites pour chaque observation ont permis de calculer une grande variété d'indices spectraux. Le NDVI fait partie des indices spectraux les plus utilisés pour

identifier la présence et la vigueur de la végétation. La Figure 7 illustre l'évolution du NDVI sur plusieurs dates pour chaque groupe d'essences. La différence temporelle est particulièrement marquée chez les feuillus, avec des valeurs nettement plus faibles pendant les mois de mai et octobre. La classe autre résineux montre une variation plus graduelle entre le mois, il faut noter que la présence de mélèzes larcin dans cette classe peut influencer à la baisse les valeurs de NDVI au printemps et à l'automne. Les trois classes restantes de résineux (épinette noire et blanche, et sapin baumier) ne montrent qu'une faible variation entre les différentes dates. Il faut noter la présence d'outliers (données aberrantes) dans chaque classe. Les différences de variation temporelle entre les classes illustrent bien l'intérêt d'utiliser plusieurs dates pour discerner différents types d'essences.

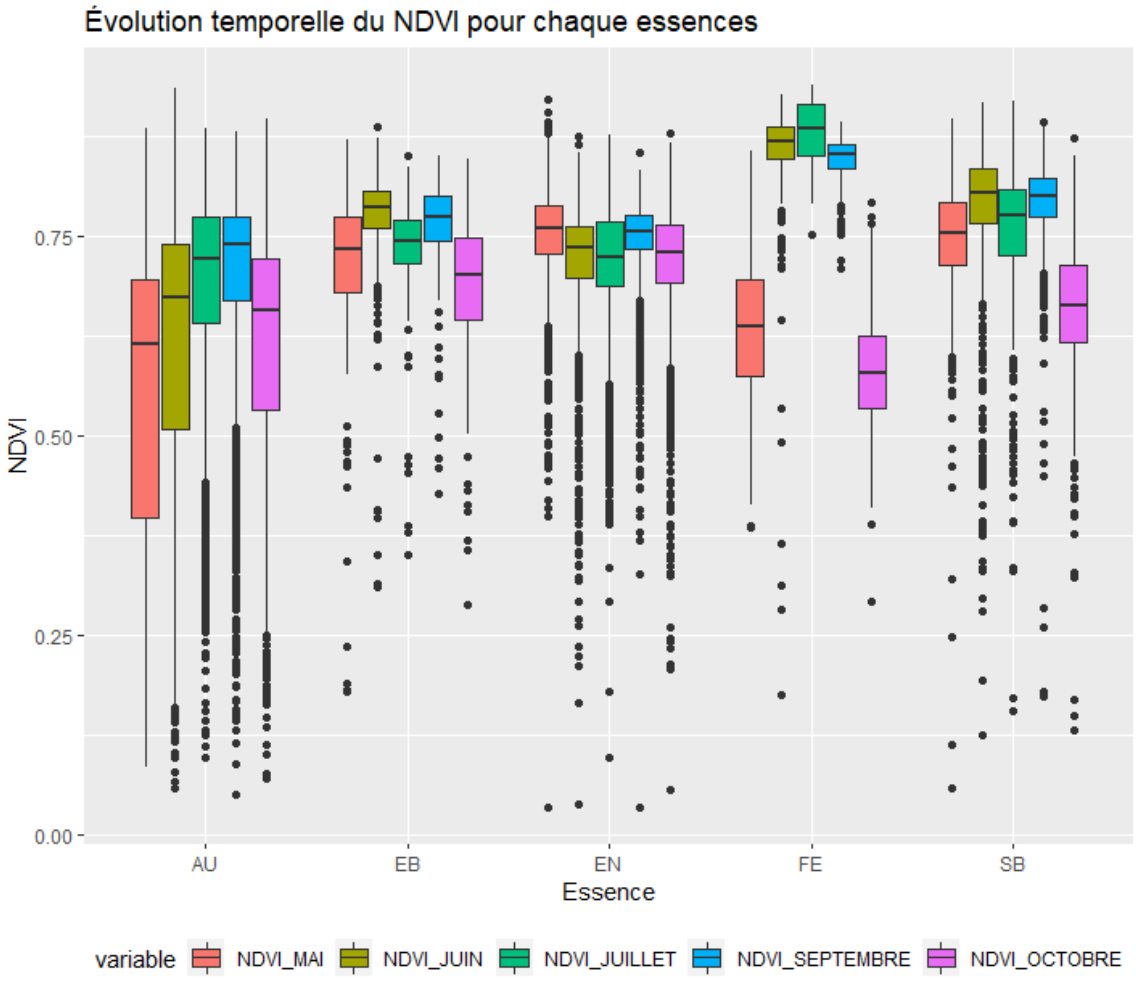


Figure 7 . Évolution temporelle de l'indice spectral NDVI selon les essences forestières d'intérêt (AU = autre résineux, EB = épinette blanche, EN = épinette noire, FE = Feuillus, SB = sapin baumier).

5.2. Modèles retenus

Plusieurs modèles random forest ont été développé avec l'ajout séquentiel de dates et le processus de sélection de variables. Les performances des modèles testés sont compilées dans

le Tableau 6. L'ajout de dates permet d'augmenter significativement les performances des modèles, celles-ci évoluant entre 0.600 et 0.877 de kappa. Le modèle utilisant cinq dates présente la meilleure performance avec 0.861 de kappa. L'application du processus de sélection de variables a permis de passer de 406 variables utilisées par le modèle à 21 variables, et une augmentation des performances à 0.877 de kappa. Considérant une utilisation opérationnelle des modèles, l'acquisition de cinq dates peut présenter un défi, particulièrement au Québec où annuellement le nombre de pixel sans nuages est le plus faible au Canada. Pour permettre plus de flexibilité d'utilisation, un modèle avec moins de dates a donc aussi été sélectionné, en supplément du modèle utilisant cinq dates. Utilisant les mois de juillet et septembre, la sélection de variable a permis de faire ressortir 38 variables d'intérêt pour ce modèle, et montre des performances de 0.833 de kappa. Bien que ces performances soient en dessous de celles du modèle utilisant cinq dates, elles restent très acceptables. L'évolution des performances en fonction du nombre de date (Figure 8) est similaire à l'évolution observée par Immitzer et al. (2019) et Grabska et al. (2019) dans un contexte similaire, ce qui confirme la pertinence des résultats obtenus.

Tableau 6 . Performances enregistrées des modèles random forest avec plusieurs combinaisons de dates et de sélection de variables.

Nombre de date	**Mois utilisé**					**Sélection variables**	**Nombre variables**	**Précision globale**	**Kappa**
	Mai	Juin	Juil.	Sept.	Oct.				
1	■					non	82	0,785	0,680
1		■				non	82	0,804	0,713
1			■			non	82	0,847	0,775
1			■			oui	21	0,846	0,774
1				■		non	82	0,801	0,707
1					■	non	82	0,735	0,600
2		■	■			non	163	0,880	0,824
2		■		■		non	163	0,852	0,783
2			■	■		non	163	0,882	0,828
2			■	■		oui	38	0,886	0,833
2	■				■	non	163	0,832	0,750
2	■		■			non	163	0,880	0,825
3		■	■	■		non	244	0,892	0,842
3		■	■	■		oui	16	0,875	0,816
3		■	■	■		oui	45	0,894	0,845
3	■		■	■		non	244	0,882	0,828
3	■		■		■	non	244	0,848	0,777
3	■	■		■		non	244	0,875	0,816
3	■	■	■			non	244	0,887	0,835
4		■	■	■	■	non	325	0,905	0,861
4		■	■	■	■	oui	44	0,887	0,835

4		non	325	0,903	0,858
4		non	325	0,889	0,837
4		non	325	0,900	0,854
4		non	325	0,892	0,841
4		oui	20	0,896	0,847
5		non	406	0,905	0,861
5		oui	21	0,918	0,877

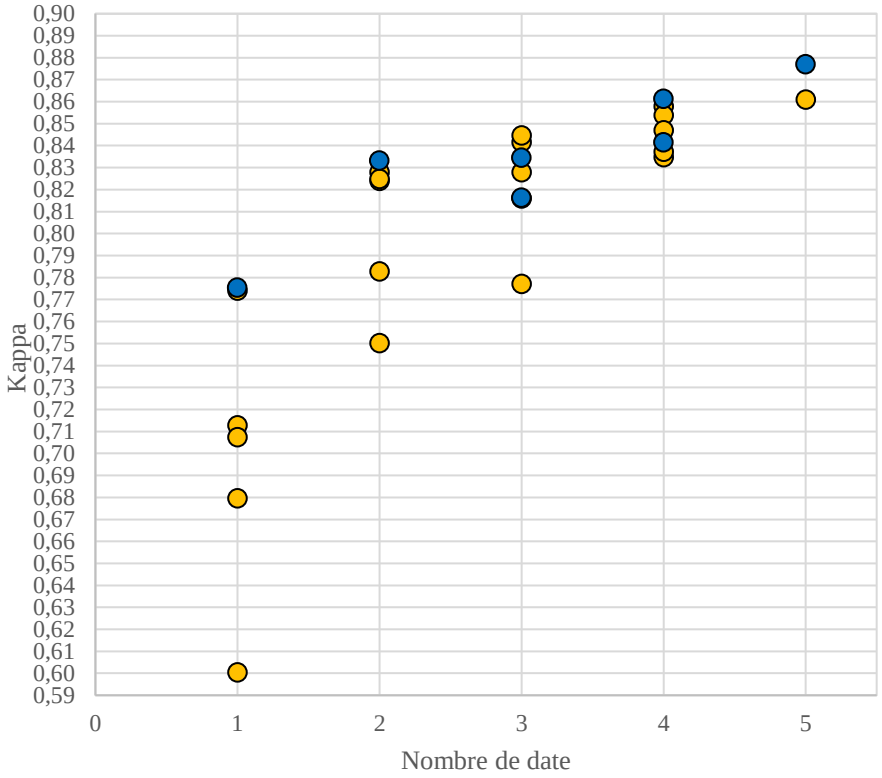


Figure 8 . Évolution des performances (Kappa de Cohen) selon le nombre de date intégrées dans le modèle random forest (JAUNE = Pas de sélection de variable; BLEU = Sélection de variables).

5.2.1. Modèle cinq dates

Les performances du modèle utilisant cinq dates et 21 variables ont été analysées en détail selon les prédictions effectuées sur les données de validation. La matrice de confusion qui en résulte permet d'identifier les classes les plus justement identifiées (Tableau 7). L'identification des feuillus est quasiment parfaite, l'épinette noire et les autres résineux montrent également un excellent succès de classification. L'épinette blanche montre un très bon score de classification, mais présente un nombre plus faible d'observations. Le sapin baumier quant à lui montre de bonnes performances, mais présente une légère confusion avec

l'épinette noire (21 sapins baumier ont été prédit par le modèle comme étant de l'épinette noire).

Tableau 7 . Matrice de confusion du modèle random forest développé avec 5 dates présentant le meilleur score de Kappa (AU = autre résineux, EB = épinette blanche, EN = épinette noire, FE = Feuillus, SB = sapin baumier).

	Référence				
Prédiction	AU	EB	EN	FE	SB
AU	357	6	8	1	11
EB	1	34	0	0	1
EN	17	4	410	0	21
FE	0	0	0	68	0
SB	6	8	4	1	104
				Précision globale	0.918
				95% IC	(0.898,0.932)
				Kappa	0.877

L'importance des variables extraite du modèle random forest est présenté dans la Figure 9. L'importance d'une variable est basée sur pourcentage de changement de l'erreur associée à un changement de la valeur de la variable. L'indice spectral TCI du mois de mai est la variable ayant le plus d'importance, suivi par la bande 6 du mois d'octobre et l'indice spectral LCCI du mois de septembre. Parmi les variables les plus importantes on retrouve une grande proportion de variables issues des mois de printemps et d'automne, montrant qu'il existe à ces périodes de plus grandes différences phénologiques entre les types d'essences qui peuvent être mesurées spectralement.

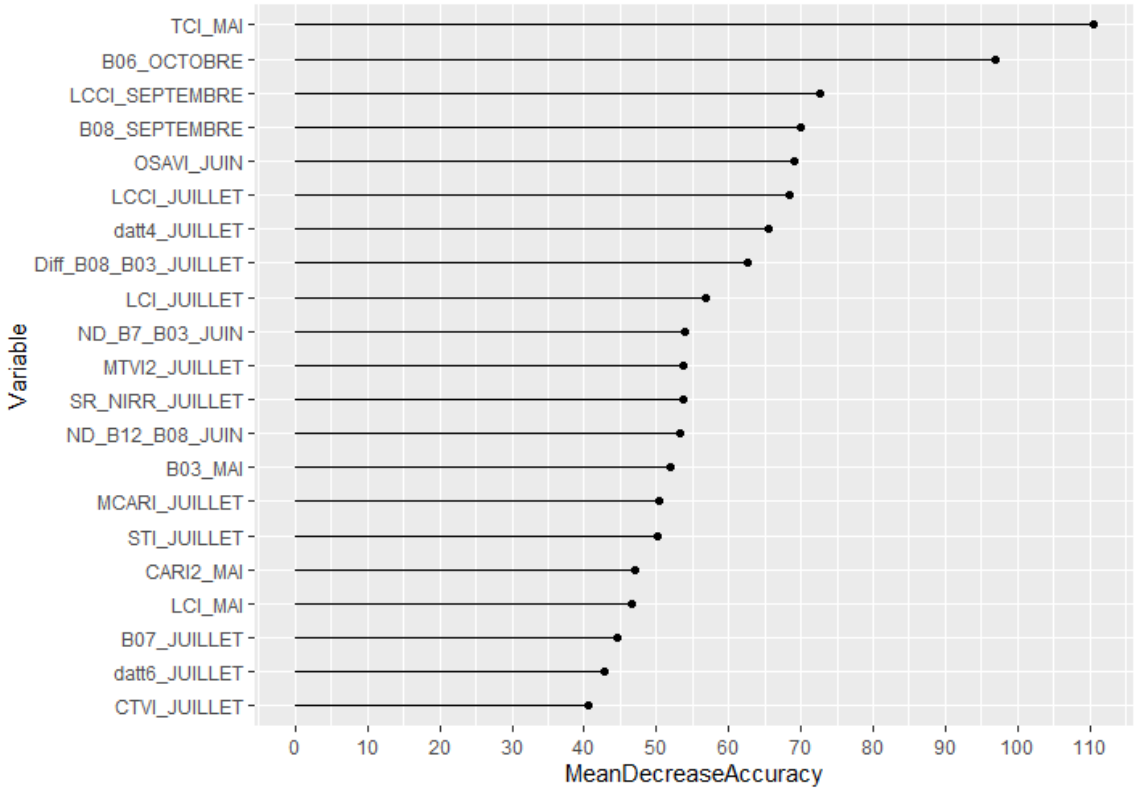


Figure 9 . Importance des variables basé sur le mean decrease accurary du modèle random forest développé à partir de cinq dates différentes.

5.2.2. Modèle deux dates.

Les performances du modèle utilisant deux dates et 38 variables ont été analysées en détail selon les prédictions effectuées sur les données de validation. La matrice de confusion qui en résulte permet d'identifier les classes les plus justement identifiées (Tableau 8). Similairement au modèle à cinq dates, l'identification des feuillus est quasiment parfaite (Tableau 7 et Tableau 8). L'épinette noire montre un bon succès de classification, mais présente quelques erreurs avec des prédictions dans « autre résineux » (Tableau 8). Les erreurs de classification concernant l'épinette blanche sont similaires avec le modèle à cinq dates, où la majorité des observations sont correctement classifiées, mais une minorité est confondue avec des sapins baumier, autre résineux et épinette noires. La classe autre résineux montre de très bonnes performances, avec la large majorité des observations correctement classifiées (Tableau 8). Le sapin baumier est en majorité correctement classifié, et une minorité d'erreurs sont distribués entre plusieurs classes similairement au modèle à cinq dates (Tableau 7 et Tableau 8).

Tableau 8 . Matrice de confusion du modèle random forest développé avec 2 dates présentant le meilleur score de Kappa (AU = autre résineux, EB = épinette blanche, EN = épinette noire, FE = Feuillus, SB = sapin baumier).

	Référence				
Prédiction	AU	EB	EN	FE	SB
AU	355	8	21	2	18
EB	4	32	0	0	6
EN	12	3	393	0	17
FE	0	0	0	65	0
SB	10	9	8	3	96
				Précision globale	0.886
				95% IC	(0.865,0. 905)
				Kappa	0.833

L'importance des variables du modèle à deux dates est illustrée dans la Figure 10. La variable avec le plus d'importance est l'indice spectral Datt1 du mois de juillet, suivi par l'indice spectral REIP2 du mois de septembre et du mois de juillet. La sélection d'un plus grand nombre de variables dans le modèle à deux dates (38 contre 21 dans le modèle a cinq dates) semble venir compenser le poids plus faible des variables. La sélection des variables a fait ressortir 23 variables du mois de juillet et 15 variables du mois de septembre. Avec les processus de sélection de variables réalisées, on constate que des variables, tel que l'indice spectral LCCI du mois de septembre, sont retrouvés dans les deux modèles, montrant la pertinence des variables sélectionnées.

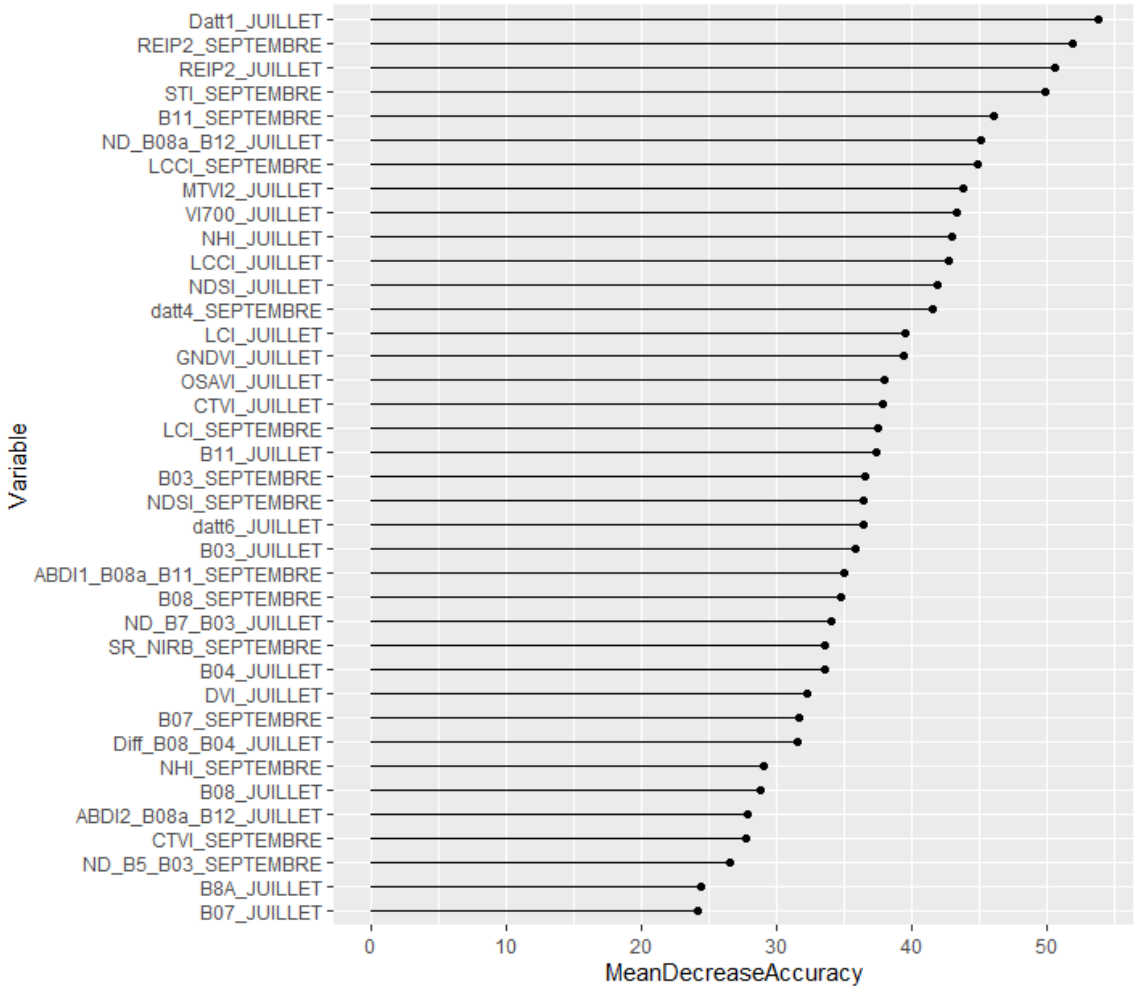


Figure 10. Importance des variables basé sur le mean decrease accurary du modèle random forest développé à partir de 2 dates différentes.

5.3. Application des modèles

5.3.1. Modèle cinq dates

L'application du modèle utilisant cinq dates à l'ensemble du territoire d'étude permet de distinguer des grands ensembles d'essences (Figure 11). Les feuillus sont nettement dominants dans la partie ouest du territoire d'étude, ce qui correspond avec les données écoforestières. Une proportion de sapins baumier plus importante peut être distinguée au sud de la ville d'Alma. L'est du territoire d'étude ainsi que la portion au nord du fjord du Saguenay est très diversifiée et présente de grands ensembles d'épinettes noires.

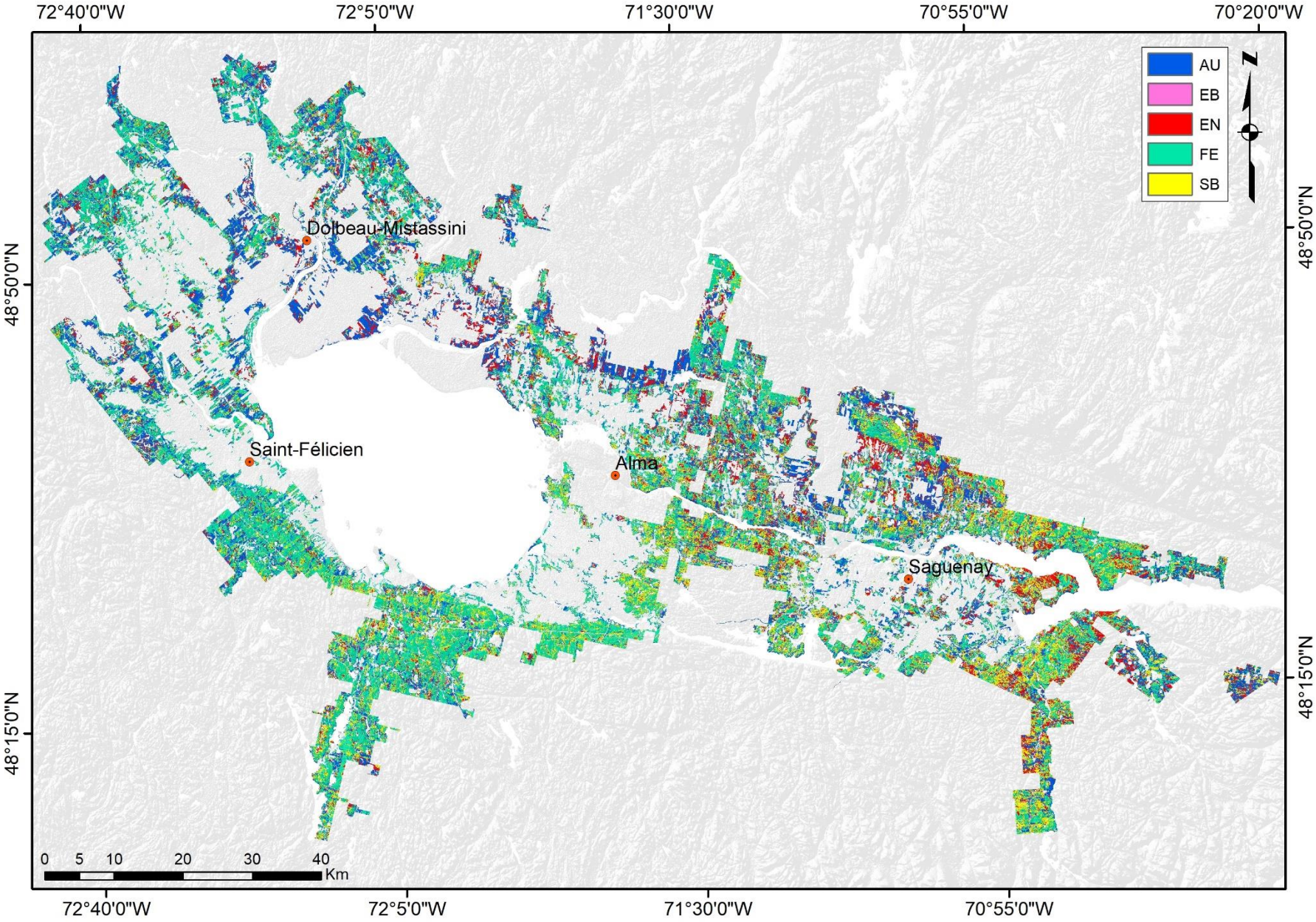


Figure 11. Cartographie des essences forestière d'intérêt pour toutes les surfaces forestières du territoire d'étude avec le modèle utilisant 5 dates (AU = autre résineux, EB = épinette blanche, EN = épinette noire, FE = Feuillus, SB = sapin baumier).

5.3.2. Modèle deux dates

L'application du modèle utilisant deux dates à l'ensemble du territoire d'étude permet de distinguer des grands ensembles d'essences similaires à ceux identifiés par le modèle utilisant cinq dates (Figure 11 et Figure 12). Les feuillus sont nettement dominants dans la partie ouest du territoire d'étude. Une proportion de sapins baumier plus importante peut être distinguée au sud de la ville d'Alma. L'est du territoire d'étude ainsi que la portion au nord du fjord du Saguenay est très diversifiée et présente de grands ensembles d'épinettes noires.

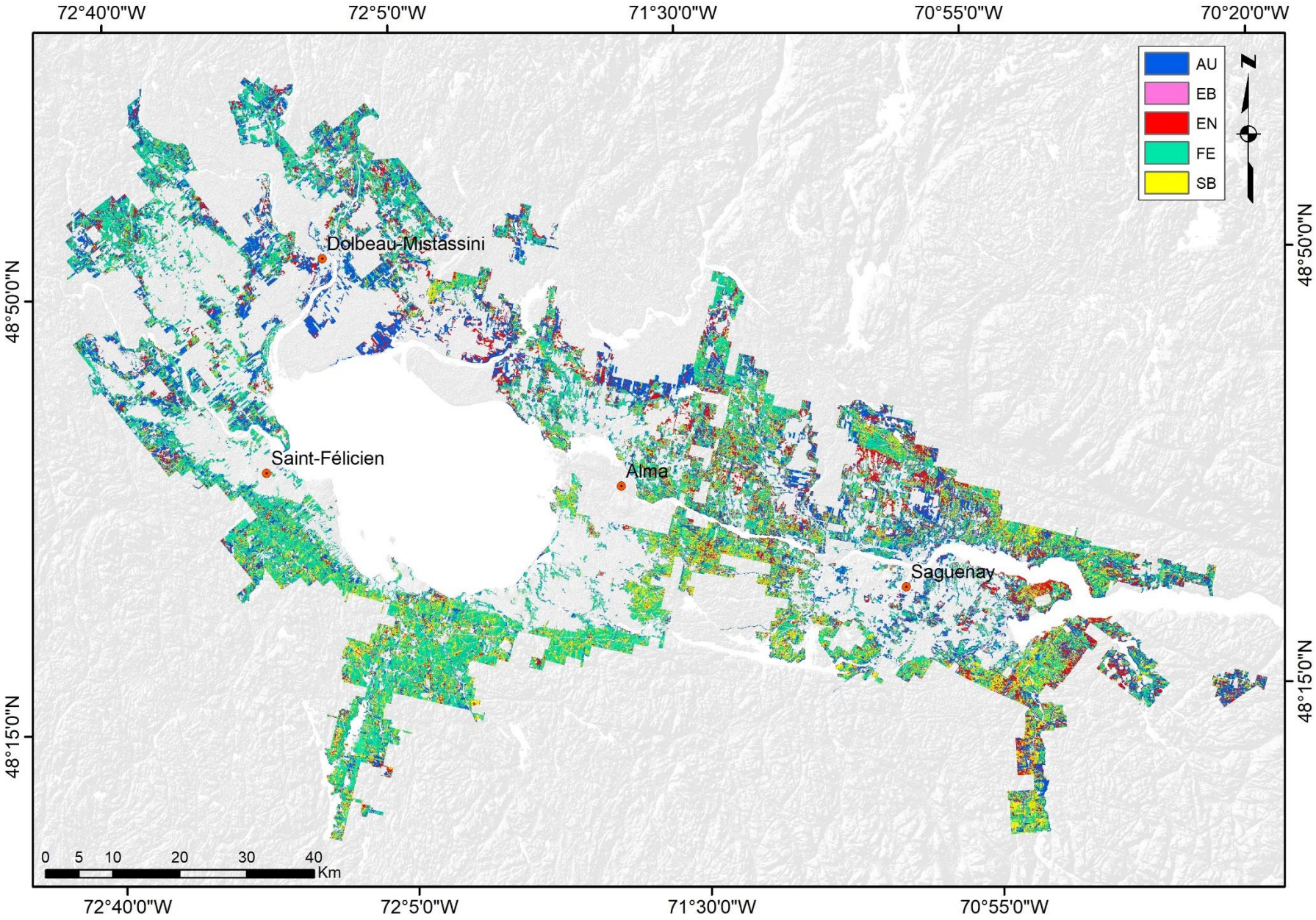


Figure 12 . Cartographie des essences forestière d'intérêt pour toutes les surfaces forestières du territoire d'étude avec le modèle utilisant 2 dates (AU = autre résineux, EB = épinette blanche, EN = épinette noire, FE = Feuillus, SB = sapin baumier).

5.3.3. Comparaisons

La Figure 13 présente une section du territoire d'intérêt où a été appliqué le modèle utilisant deux dates et le modèle utilisant cinq dates. Il est possible de distinguer dans les classifications une concordance avec les grands ensembles de végétation retrouvés dans la carte écoforestière.

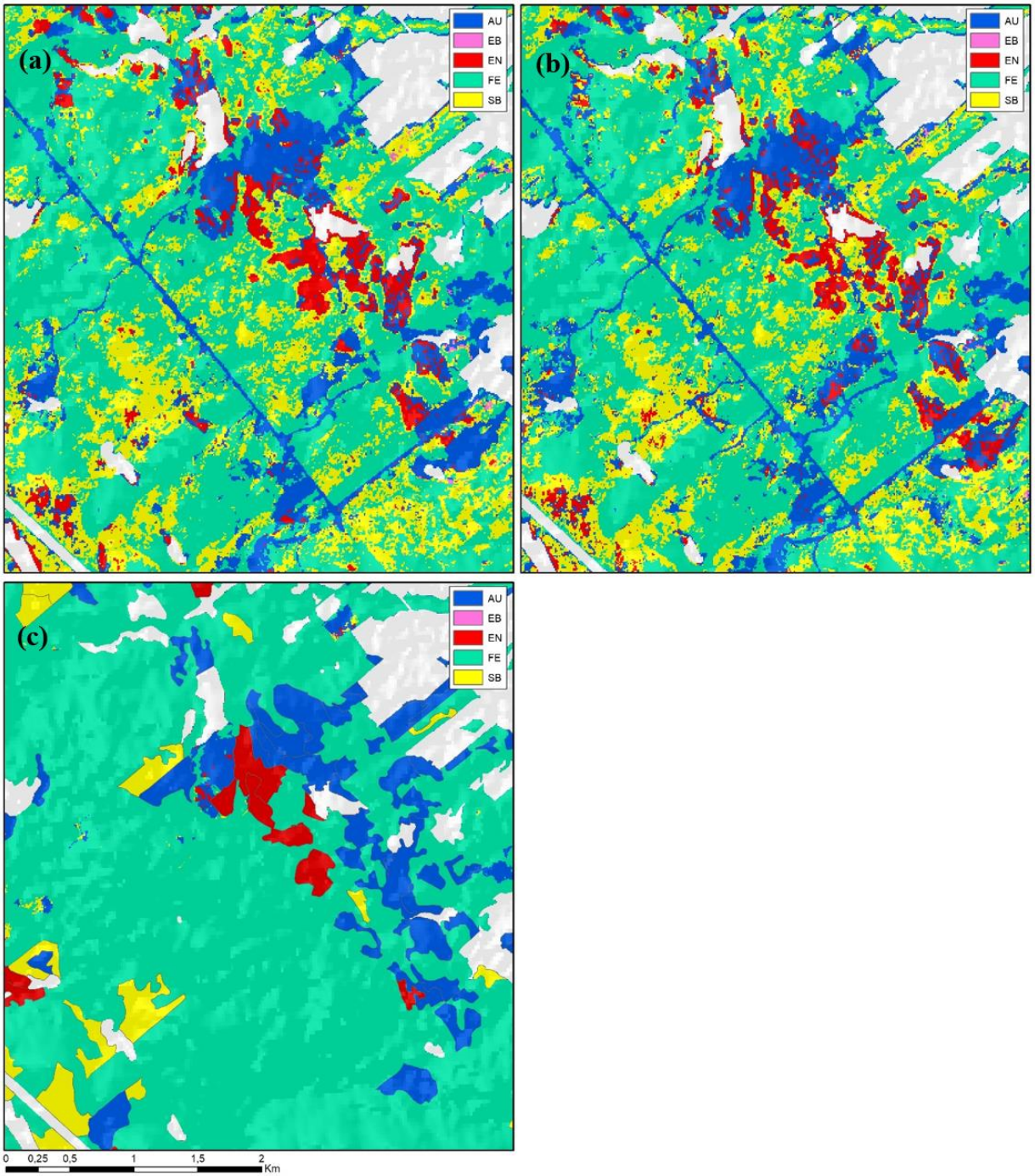


Figure 13. Comparaison entre les résultats de classification des modèles à cinq dates (a), deux dates (b) et la carte écoforestière (c), sur une portion du territoire d'intérêt.

La Figure 14 illustre quelques exemples de comparaisons entre les points photo-interprétés et les classifications. Les deux modèles montrent une très bonne capacité de discernement entre les essences et sont capables d'identifier des pixels isolés d'essences pures.

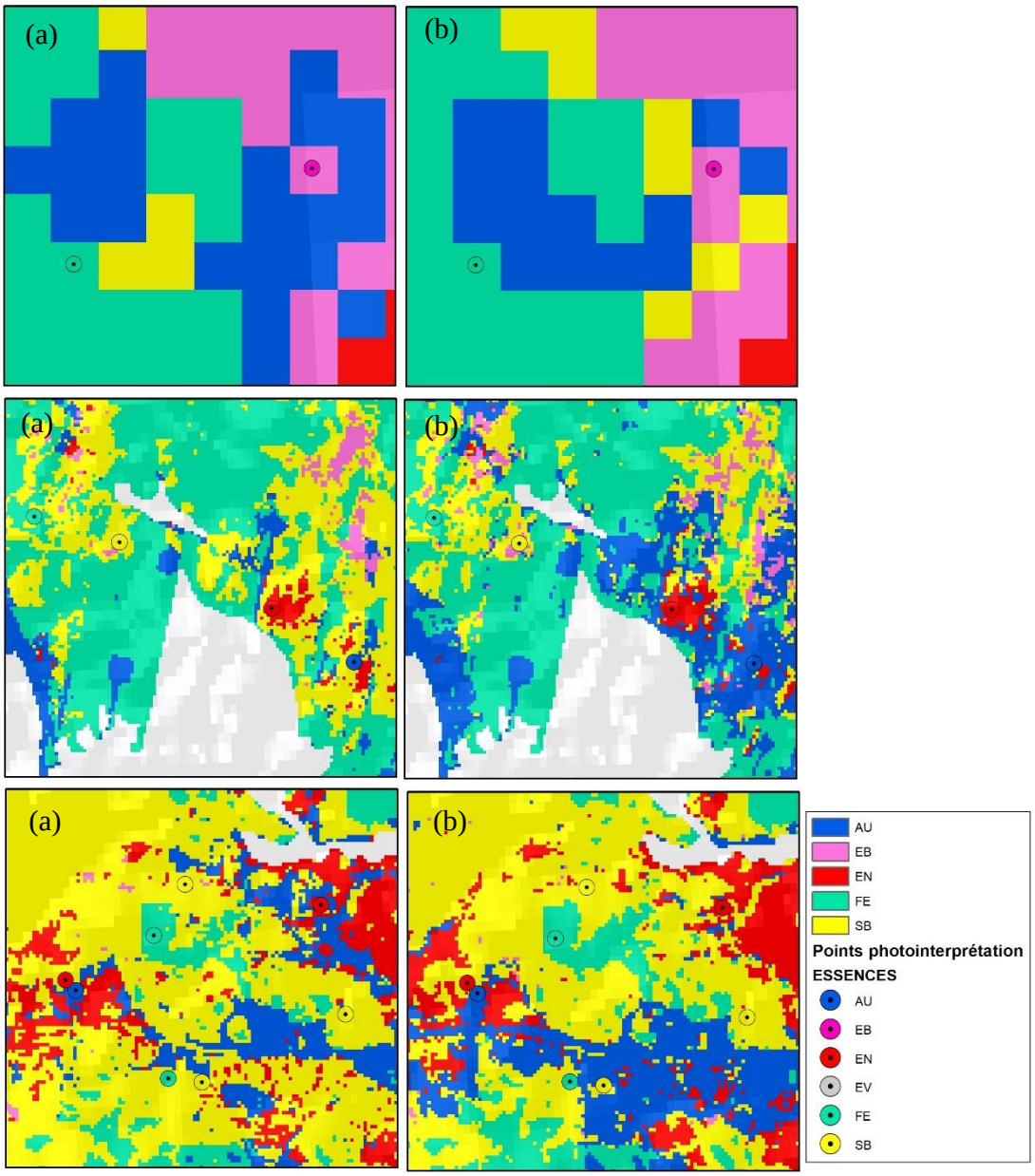


Figure 14. Exemple de classifications des modèles deux dates (a) et cinq dates (b) en comparaison avec plusieurs points photo-interprétés.

5.3.4. Classification de l'épinette de Norvège

Bien que n'ayant pas pu être inclue dans le développement du modèle à cause d'une trop faible occurrence dans le territoire d'étude, l'épinette de Norvège reste une essence d'intérêt lors des prises de décisions. La classification des pixels identifiées par photo-interprétation comme étant de l'épinette de Norvège a été extraite (Figure 15). Une grande majorité des

pixels d'épinette de Norvège ont été classifiés en tant qu'épinette blanche. Peu de différences existent entre les deux modèles, avec cependant la présence d'une observation classée en tant que sapin baumier pour le modèle utilisant 2 dates.

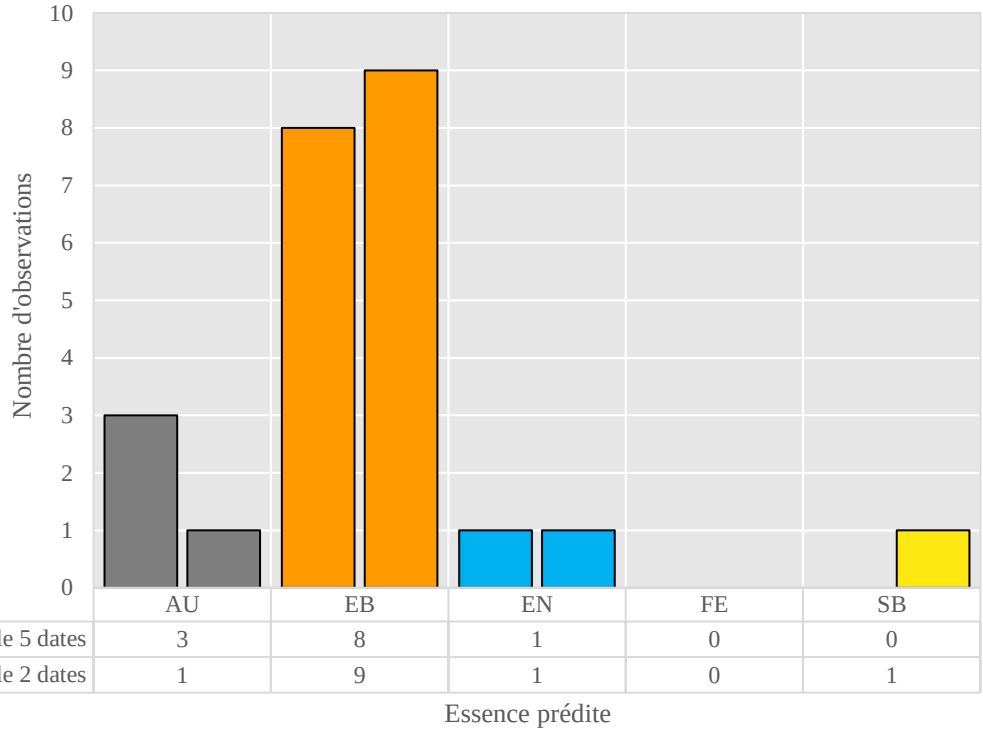


Figure 15 . Classification des 12 épinettes de Norvège identifiées par photo-interprétation (AU = autre résineux, EB = épinette blanche, EN = épinette noire, FE = Feuillus, SB = sapin baumier).

6. Discussion

La méthodologie développée a permis d'extraire et d'identifier les variables pertinentes pour classifier un vaste territoire. Les deux modèles sélectionnés montrent d'excellentes performances de classification pour les essences forestières d'intérêt.

6.1. Limites

Bien que les modèles soient performants, certaines limites ont pu être identifiées. Le lidar était pressenti comme une donnée d'intérêt pour identifier les essences forestières et a déjà été utilisé dans ce contexte (Varin *et al.*, 2020). Cependant, l'absence de feuilles dans une majorité du territoire d'étude a rendu inutilisable cette donnée.

La nécessité d'avoir cinq dates pour obtenir les meilleures performances peut être une limitation avec la présence de nuages dans les images Sentinel-2. Idéalement toutes les dates seraient issues d'une même année, cependant la disponibilité de cinq mois présentant des

acquisitions sans nuages est rare, ce qui oblige de décaler d'une ou plusieurs années pour obtenir l'information nécessaire à la cartographie d'un secteur.

Un traitement préalable a été effectué pour retirer du territoire d'étude les zones anthropiques, agricoles, humide, les coupes récentes ainsi que les routes, cependant le territoire d'étude comprends beaucoup de terres privées dans lesquels les aménagements ne sont pas toujours répertoriés. Dans cette situation, les modèles ont effectué des prédictions d'essences forestières à des emplacement ne contenant pas de forêt. Un post-traitement est donc nécessaire pour identifier des structures et aménagements présents dans une zone d'intérêt.

Initialement les arbres morts devaient être intégrés dans le modèle, cependant il y avait une trop faible occurrence dans le territoire d'étude ainsi qu'une faible distinction avec les arbres sans feuilles présents dans les images photo-interprétées, c'est pourquoi ils ont été retirés de l'analyse. Lors de la photo-interprétation, très peu de pixels contenant uniquement des arbres morts ont été identifiés. De ce fait, l'abandon de cette classe n'amène qu'une faible limitation à la capacité du modèle à identifier l'essence dominante présente dans un pixel.

Très peu d'observations d'épinette de Norvège ont été identifiées lors de la photo interprétation, causé par une faible occurrence dans le territoire d'étude ainsi qu'une forte présence de terres privées sur lesquelles les aménagements forestiers ne sont pas systématiquement rapportés jusque dans les bases de données à référence spatiale. Ce faible nombre d'observations n'a pas permis le développement d'un modèle prédictif incluant cette essence forestière.

Les modèles ont montré une excellente capacité à correctement classifier des peuplements purs, cependant aucune mesure de performances n'a été réalisé pour identifier l'effet d'un mélange d'essences au sein d'un même pixel. La proportion minimale pour considérer un pixel comme étant pur, selon le modèle, peut être variable et peut induire des erreurs de classification. Cependant, des comparaisons avec les données de la carte écoforestière semblent montrer que même dans un peuplement mixte, l'essence majoritaire est attribuée au pixel.

6.2. Pistes d'amélioration

La méthode utilisée s'est montrée très performance pour identifier des essences d'intérêt. Le regroupement en grandes catégories de plusieurs essences d'arbres a permis d'éviter la confusion avec les classes d'intérêt. Cependant certaines classes présentent un faible nombre d'observations, avec notamment l'épinette blanche qui est à la limite nord de son aire de répartition. La sélection du territoire d'étude a été contraint par la disponibilité en données lidar, cependant avec les performances enregistrées lors du développement de la méthode, l'utilisation de l'imagerie seule s'est avérée être suffisante, ce qui simplifie la gestion des données. L'utilisation de plusieurs territoires d'études, sélectionnés pour leur abondance de peuplements purs d'essences d'intérêt, permettrait d'extraire une grande quantité d'observations. Ces données permettraient d'obtenir un modèle encore plus robuste et apte à différencier des essences dans un large éventail de conditions environnementales et de mélanges d'essences. Le processus de création de modèle ne permettant pas d'ajouter de

nouvelles essences sans un nouvel entrainement, il serait pertinent d'appliquer cette amélioration si l'occasion se présente.

Avec la progression des acquisitions lidar du Québec méridional, il serait intéressant d'utiliser à nouveau cette donnée pour augmenter la robustesse du modèle. L'utilisation du lidar permettrait également d'avoir des informations sur la densité d'arbres présents, donnant ainsi des renseignements supplémentaires dans le processus de sélection des aires admissibles. Néanmoins, une couverture lidar avec feuilles est à privilégier.

7. Références

Alonzo, M., Bookhagen, B. et Roberts, D. A. 2014. Urban tree species mapping using hyperspectral and lidar data fusion. Remote Sensing of Environment, vol. 148, pp. 70-83.

Bouchard, M. 2011. Le rôle des feux et des épidémies de tordeuse des bourgeons de l'épinette noire en forêt mixte.

Boucher, D., De Grandpré, L., Kneeshaw, D., St-Onge, B., Ruel, J. C., Waldron, K. et Lussier, J. M. 2015. Effects of 80 years of forest management on landscape structure and pattern in the eastern Canadian boreal forest. Landscape Ecology, vol. 30, n° 10, pp. 1913-1929.

Breiman, L. 2001. Random forests. Machine learning, vol. 45, pp. 5-32.

Chance, E. W., Cobourn, K. M., Thomas, V. A., Dawson, B. C. et Flores, A. N. 2017. Identifying irrigated areas in the Snake River Plain, Idaho: Evaluating performance across composting algorithms, spectral indices, and sensors. Remote Sensing, vol. 9, n° 6, pp. 1-23.

Chuvieco, E., Martín, M. P. et Palacios, A. 2002. Assessment of different spectral indices in the red-near-infrared spectral domain for burned land discrimination. International Journal of Remote Sensing, vol. 23, n° 23, pp. 5103-5110.

Cohen, J. 1960. A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement, vol. 20, n° 1, pp. 37-46.

Dixit, A., Goswami, A. et Jain, S. 2019. Development and evaluation of a new « Snow Water Index (SWI) » for accurate snow cover delineation. Remote Sensing, vol. 11, n° 23.

Ermolaev, N. R., Yudin, S. A., Belobrov, V. P. et Dridiger, V. K. 2020. Using spectral indices for interpretation of plant residues under different tillage systems Using spectral indices for interpretation of plant residues under different tillage systems.

Fuentealba, A., Bauce, É. et Dupont, A. 2015. Bacillus thuringiensis efficacy in reducing spruce budworm damage as affected by host tree species. Journal of Pest Science, vol. 88, n° 3, pp. 593-603.

Gagnon, R. et Morin, H. 2001. Les forêts d'épinette noire du Québec : dynamique, perturbations et biodiversité. Le naturaliste canadien, vol. 125, n° Automne 2001, pp. 26-35.

Gamon, J. A., Field, C. B., Goulden, M. L., Griffin, K. L., Hartley, A. E., Joel, G., Penuelas, J. et Valentini, R. 1995. Relationships Between NDVI, Canopy Structure, and Photosynthesis in Three Californian Vegetation Types. Ecological Applications, vol. 5, n° 1, pp. 28-41.

Genuer, R., Poggi, J.-M. et Tuleau-Malot, C. 2019. VSURF: Variable Selection Using Random Forests.

Girard, F., Payette, S. et Gagnon, R. 2008. Rapid expansion of lichen woodlands within the closed-crown boreal forest zone over the last 50 years caused by stand disturbances in eastern Canada. Journal of Biogeography, vol. 35, n° 3, pp. 529-537.

Grabska, E., Hostert, P., Pflugmacher, D. et Ostapowicz, K. 2019. Forest stand species mapping using the sentinel-2 time series. Remote Sensing, vol. 11, n° 10, pp. 1-24.

Hagen-Zanker, A. 2006. Map comparison methods that simultaneously address overlap and structure. Journal of Geographical Systems, vol. 8, n° 2, pp. 165-185.

Immitzer, M., Neuwirth, M., Böck, S., Brenner, H., Vuolo, F. et Atzberger, C. 2019. Optimal input features for tree species classification in Central Europe based on multi-temporal Sentinel-2 data. Remote Sensing, vol. 11, n° 22.

Kobayashi, N., Tani, H., Wang, X. et Sonobe, R. 2020. Crop classification using spectral indices derived from Sentinel-2A imagery. Journal of Information and Telecommunication, vol. 4, n° 1, pp. 67-90.

Leboeuf, A., Pomerleau, I. et Québec (Province). Direction des inventaires forestiers 2015. Projet d'acquisition de données par le capteur LiDAR à l'échelle provinciale : analyse des retombées et recommandations.

Liaw, A. et Wiener, M. 2002. Classification and Regression by randomForest. R News, vol. 2, n° 3, pp. 18-22.

Liu, H., Xu, L., Ding, J. et Deng, X. 2010. A new angle-based spectral index and its application in drought monitoring. 2010 Symposium on Photonics and Optoelectronic, SOPO 2010 - Proceedings.

Marrs, J. et Ni-Meister, W. 2019. Machine learning techniques for tree species classification using co-registered LiDAR and hyperspectral data. Remote Sensing, vol. 11, n° 7, pp. 1-18.

MFFP, (Ministère des Forêts de la Faune et des Parcs) 2018. Cartographie du 5e inventaire écoforestier du Québec méridional — Méthodes et données associées, pp. 111 p.

Monserud, R. A. et Leemans, R. 1992. Comparing global vegetation maps with the Kappa statistic. Ecological Modelling, vol. 62, n° 4, pp. 275-293.

Morcillo-Pallarés, P., Rivera-Caicedo, J. P., Belda, S., De Grave, C., Burriel, H., Moreno, J. et Verrelst, J. 2019. Quantifying the robustness of vegetation indices through global sensitivity analysis of homogeneous and forest leaf-canopy radiative transfer models. Remote Sensing, vol. 11, n° 20, pp. 1-23.

Mustafa, D. M. T., Hassoon, D. K. I., Hussain, D. H. M. et Abd, M. H. 2017. USING WATER INDICES (NDWI, MNDWI, NDMI, WRI AND AWEI) TO DETECT PHYSICAL AND CHEMICAL PARAMETERS BY APPLY REMOTE SENSING AND GIS TECHNIQUES.

Pimstein, A., Eitel, J. U. H., Long, D. S., Mufradi, I., Karnieli, A. et Bonfil, D. J. 2009. A spectral index to monitor the head-emergence of wheat in semi-arid conditions. Field Crops Research, vol. 111, n° 3, pp. 218-225.

Probst, P., Wright, M. N. et Boulesteix, A. L. 2019. Hyperparameters and tuning strategies for random forest. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, vol. 9, n° 3, pp. 1-19.

Quemada, M. et Daughtry, C. S. T. 2016. Spectral indices to improve crop residue cover estimation under varying moisture conditions. Remote Sensing, vol. 8, n° 8.

R Core Team 2020. R: A Language and Environment for Statistical Computing. Vienna, Austria.

Shojanoori, R. et Shafri, H. Z. M. 2016. Review on the use of remote sensing for urban forest monitoring. Arboriculture & Urban Forestry, vol. 42, n° 6, pp. 400-417.

Varin, M., Chalghaf, B., Joanisse, G. et Dubois, A.-M. 2020. Classification d'essences forestières par une approche de modélisation de type « Random Forest », à partir du lidar et de Sentinel-2 - forêt de Black Brooke, NouveauBrunswick. Centre d'enseignement et de recherche en foresterie de Sainte-Foy inc. (CERFO). Rapport 2020-01, 34 p.

White, J. C., Wulder, M. A., Vastaranta, M., Coops, N. C., Pitt, D. et Woods, M. 2013. The utility of image-based point clouds for forest inventory: A comparison with airborne laser scanning. Forests, vol. 4, n° 3, pp. 518-536.

Wulder, M. A., White, J. C., Loveland, T. R., Woodcock, C. E., Belward, A. S., Cohen, W. B., Fosnight, E. A., Shaw, J., Masek, J. G. et Roy, D. P. 2016. The global Landsat archive: Status, consolidation, and direction. Remote Sensing of Environment, vol. 185, pp. 271-283.

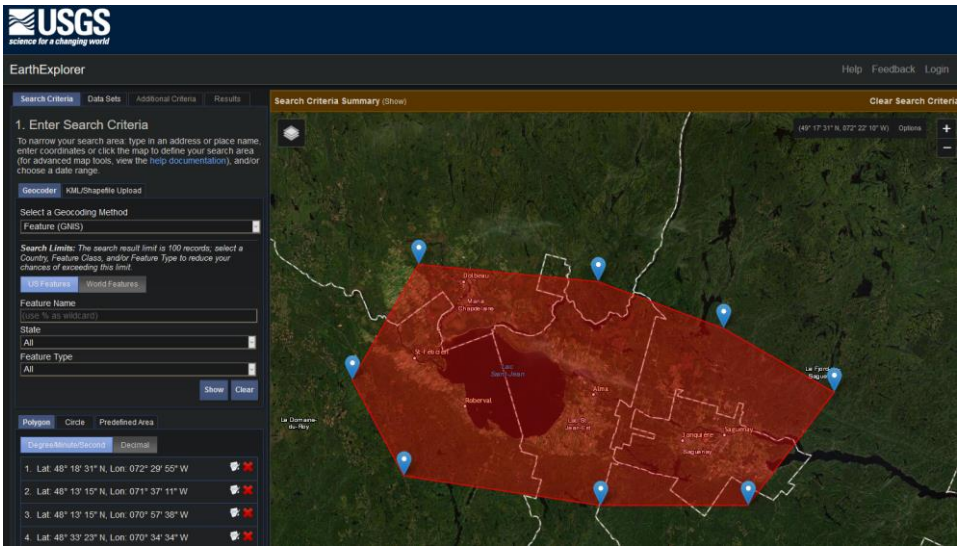
Xue, J. et Su, B. 2017. Significant remote sensing vegetation indices: A review of developments and applications. Journal of Sensors. Hindawi Limited.

8. Annexe : Guide d'utilisation

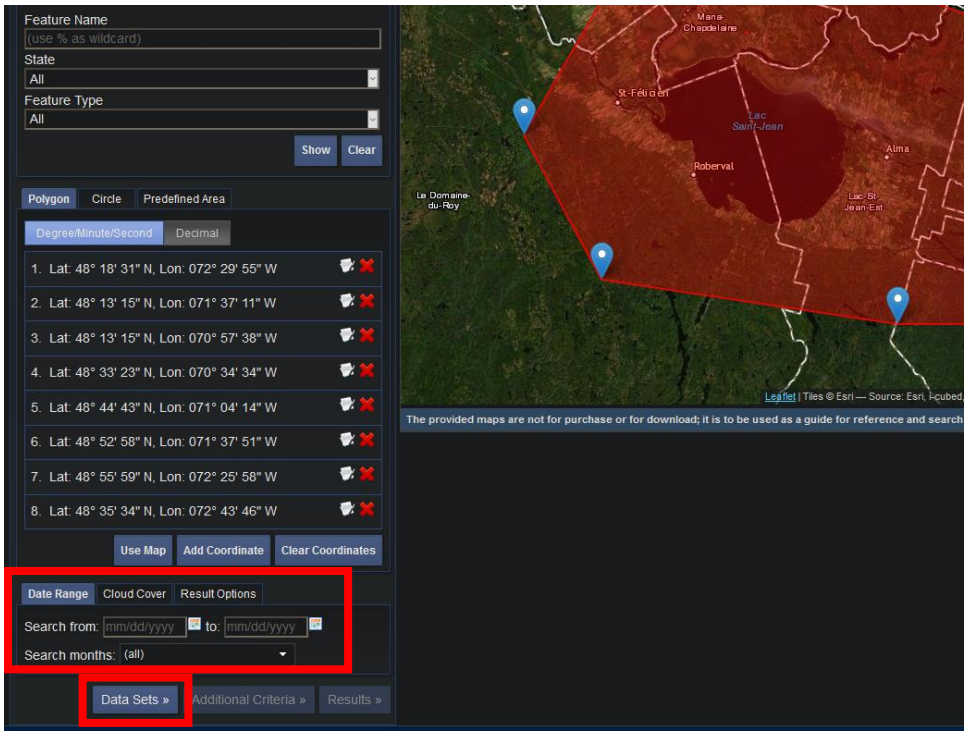
Étape 1 : Téléchargement des images Sentinel-2

Téléchargement avec EarthExplore (https://earthexplorer.usgs.gov/)

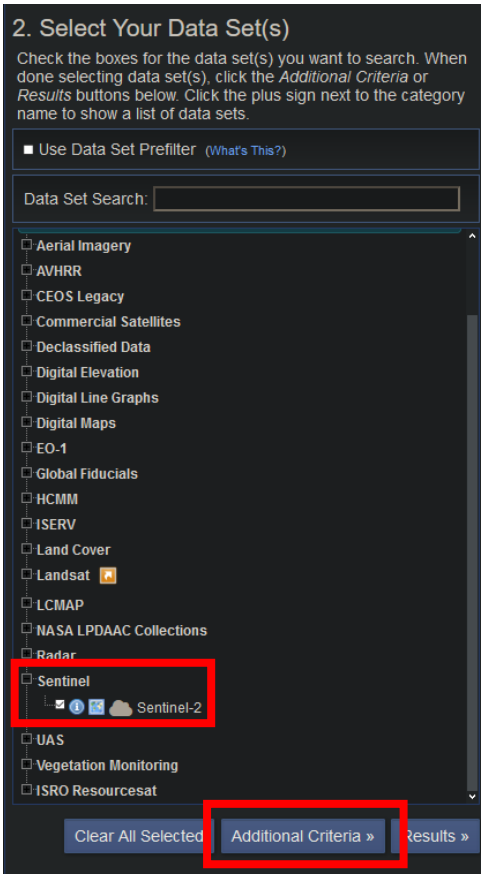
1. Créer un compte/se connecter
2. Définir un territoire d'intérêt



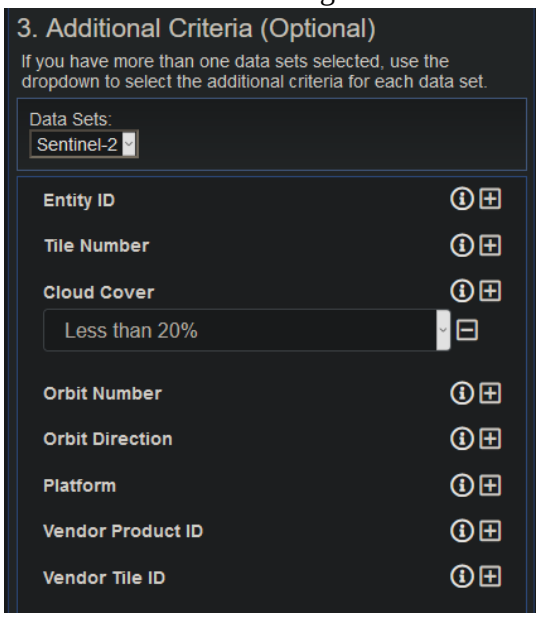
3. Définir un intervalle de dates pour la recherche d'images, puis cliquer sur « data sets » :



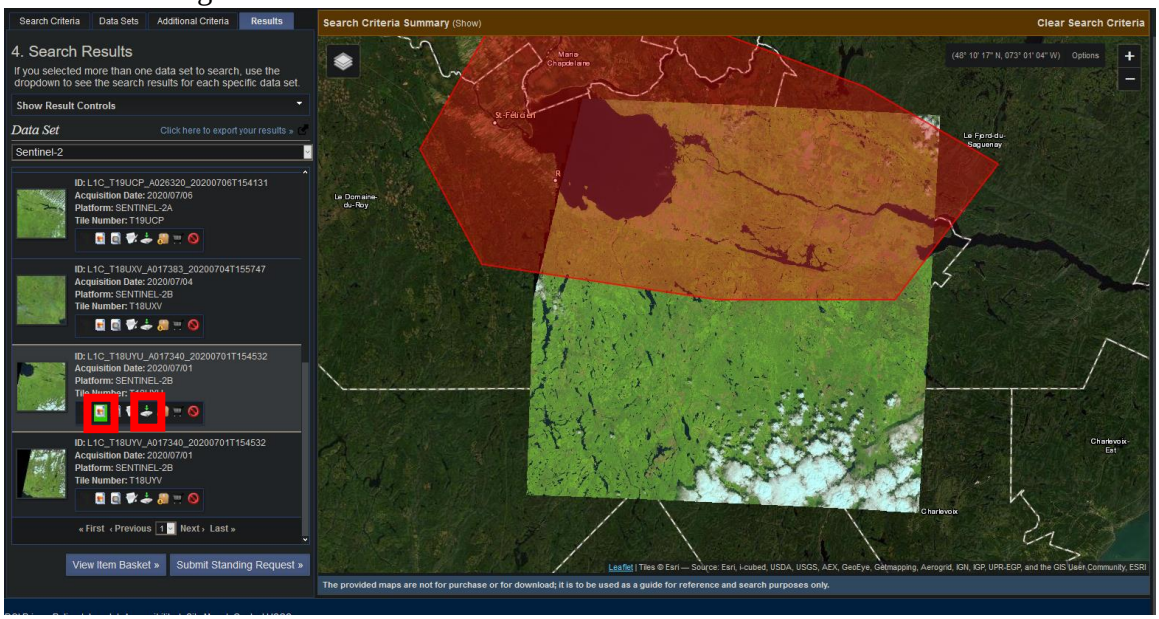
4. Dans la liste, dérouler « Sentinel » puis cocher « Sentinel-2 ». Cliquer ensuite sur « Additional Criteria »

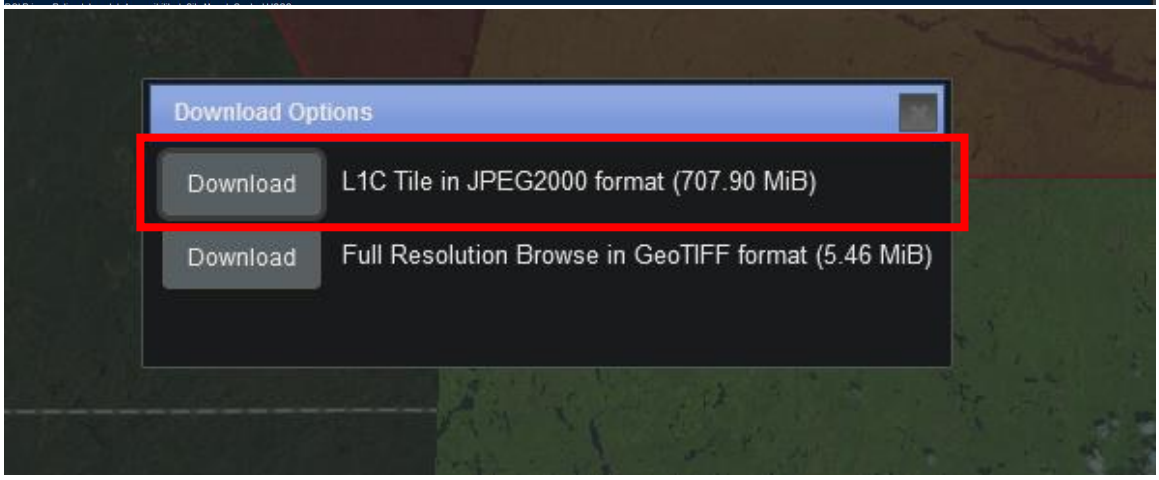


5. Définir un seuil minimal de couverture nuageuse :



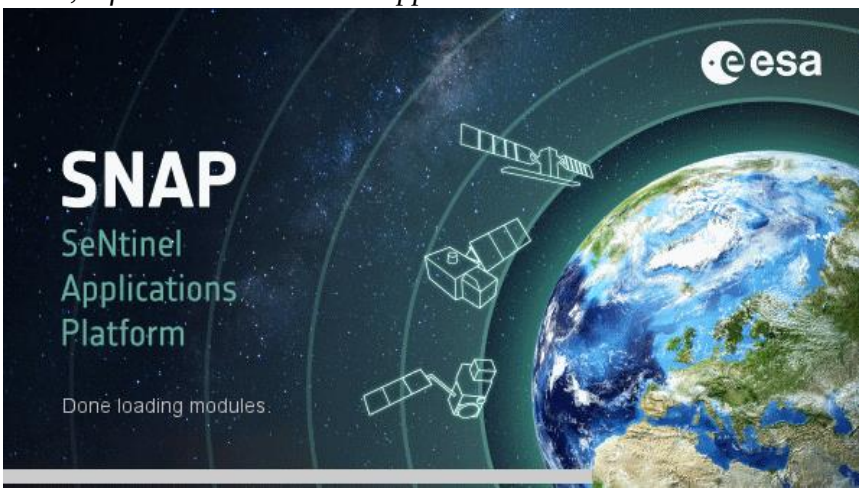
6. Identifier des images convenables pour renseigner la zone d'intérêt, puis les télécharger dans le bon format :





Étape 2 : Corrections atmosphériques dans SNAP

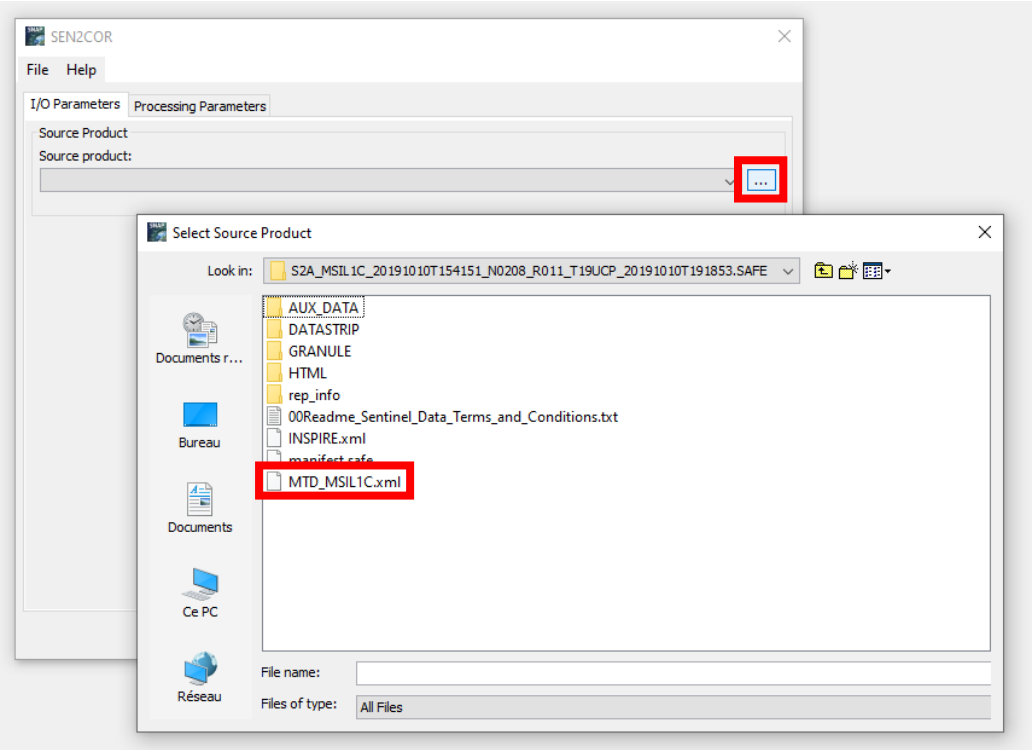
Si les images Sentinel-2 ont été téléchargées au niveau « L2A », cette étape n'est pas nécessaire. Sinon, il faut installer les deux applications suivantes.



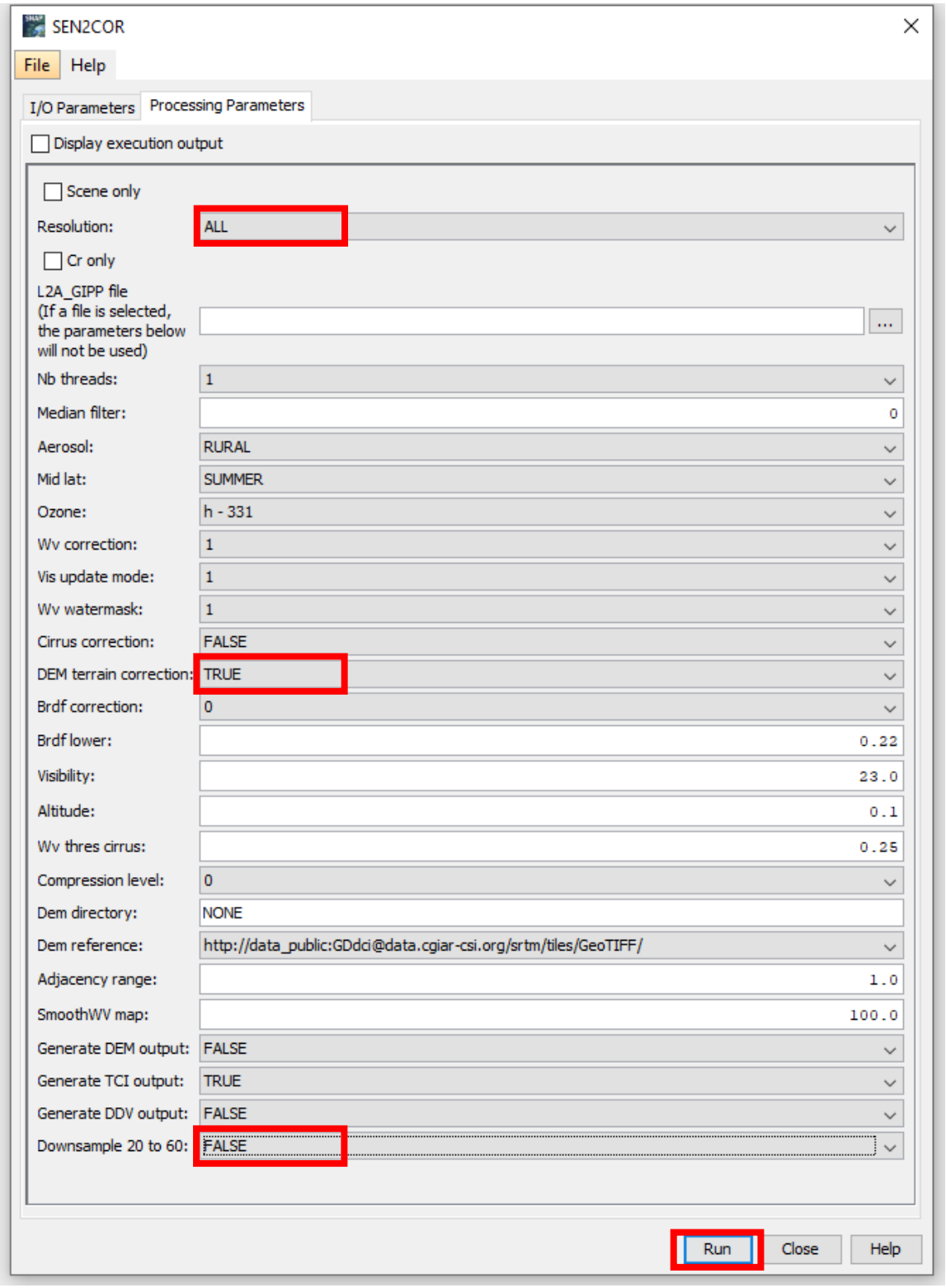
Logiciel SNAP : https://step.esa.int/main/download/snap-download/
Plugin Sen2Cor v2.8: http://step.esa.int/main/snap-supported-plugins/sen2cor/sen2cor_v2-8/

L'étape de corrections atmosphérique doit être effectuée image par image.

1. Ouvrir le plugin Sen2Core depuis les menus : Optical > Thematic Land Processing > Sen2Cor Processor > Sen2Core280
2. Ouvrir une image Sentinel-2 :



3. Dans l'onglet « Processing parameters », changer les paramètres suivants, puis cliquer sur « Run » :



Étape 3 : Territoire d'intérêt

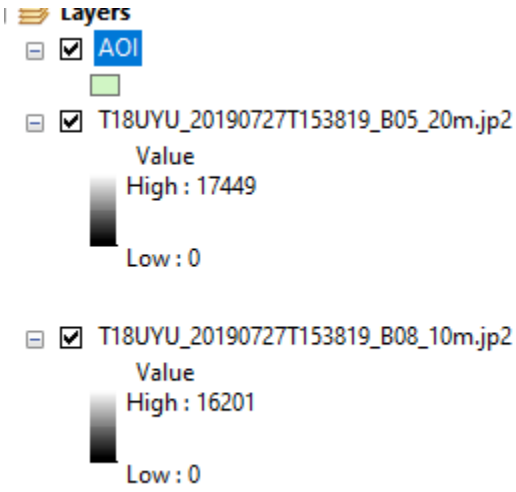
Afin de découper les images Sentinel-2 avec le territoire d'intérêt, trois fichiers sont nécessaires :

- Territoire d'intérêt en .shp
- Territoire d'intérêt, raster de 10 m de résolution, avec une étendue (*extent*) dans la projection finale.
- Territoire d'intérêt, raster de 20 m de résolution, avec une étendue (*extent*) dans la projection finale.

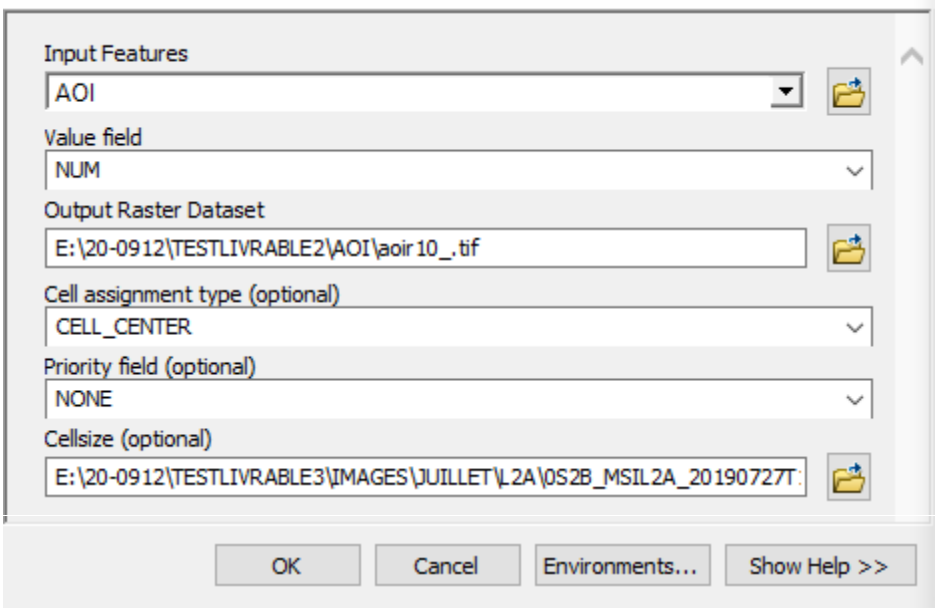
Création des rasters du territoire d'intérêt :

Fichiers nécessaires :

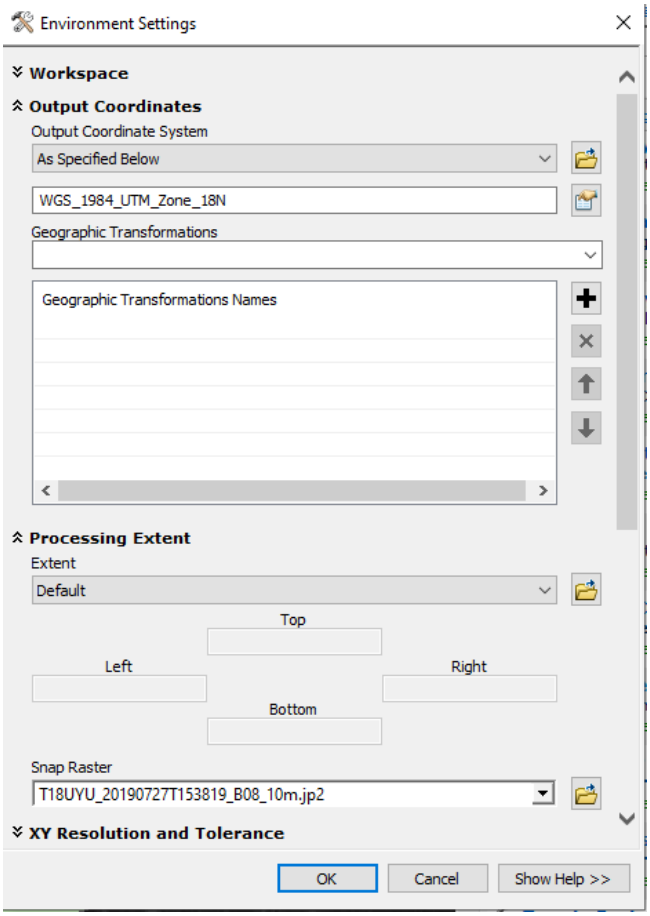
- .shp du territoire d'intérêt.
- Bande spectrale Sentinel-2 à 10 m :
  ex. :"*imageSentinel2*.SAFE\GRANULE\L2A_*nomImageSentinel2*\IMG_DATA\R10m*nomImageSentinel2* _B08_10m.jp2"
- Bande spectrale Sentinel-2 à 20m :
  ex. :"*imageSentinel2*.SAFE\GRANULE\L2A_*nomImageSentinel2*\IMG_DATA\R20m*nomImageSentinel2* _B05_20m.jp2"



Dans ArcMap, avec l'outil *Polygon To Raster*, créer les rasters du territoire d'intérêt pour chaque résolutions (10 et 20 m) avec les paramètres suivants :

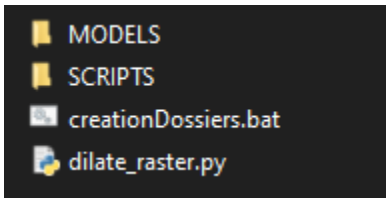


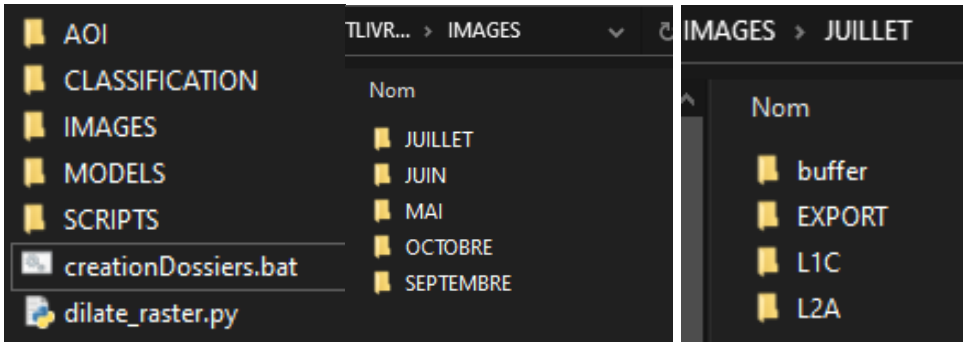
Pour le paramètre « Value Field », le .shp de la zone d'intérêt doit contenir un champ numérique, *NUM* dans l'exemple. La valeur du champ n'est pas importante. Si aucun champ numérique existe, créer un champ avec la valeur 1 pour chaque polygone par exemple. Il est important de remplir le paramètre *snap* avec le bon raster utilisé.



Étape 4 : Préparation du dossier de travail

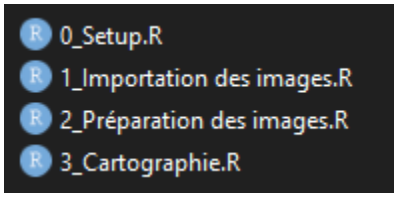
Vous pouvez commencer par installer R (https://cran.r-project.org/bin/windows/base/R-4.0.4-win.exe). Afin de pouvoir utiliser les scripts R, le dossier de travail doit être agencé avec une architecture de dossiers prédéfinie. Le fichier creationDossiers.bat permet de créer cette architecture. Il suffit de double cliquer dessus pour créer l'arborescence.





Les images Sentinel-2 peuvent être placées dans les dates leur correspondant, dans le dossier « L2A ». Le .shp du territoire d'intérêt ainsi que les rasters précédemment créés peuvent être placés dans le dossier « AOI ».

Étape 5 : Gestion des images et classification dans R



0_Setup.R : Chargement des packages et définition du dossier de travail.

Première utilisation :

- Installer python (https://docs.anaconda.com/anaconda/install/)
- Utiliser le script R pour installer les packages
- Utiliser la commande : source_python("./dilate_raster.py")

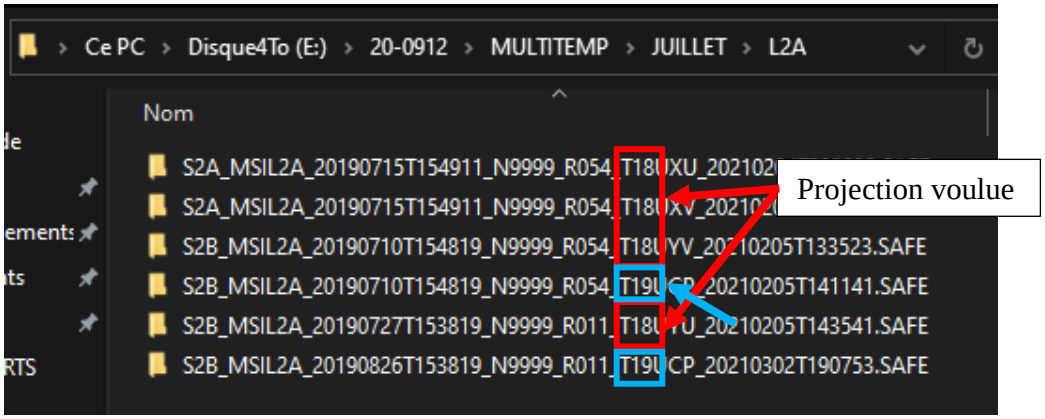
- Accepter l'installation de mini-conda (il est normal d'avoir un message d'erreur après cette commande, car les packages python ne sont pas installés).
- Ouvrir une invite de commande anaconda prompt (R-MINI)
- Entrer les commandes suivantes :
  conda activate r-reticulate
  conda install -c conda-forge opencv rasterio

1_Importation des images.R : Importation des images Sentinel-2 dans R pour chaque date. Le paramètre à modifier est le nombre d'images à utiliser dans la classification en fonction du modèle choisi (2 ou 5 images).

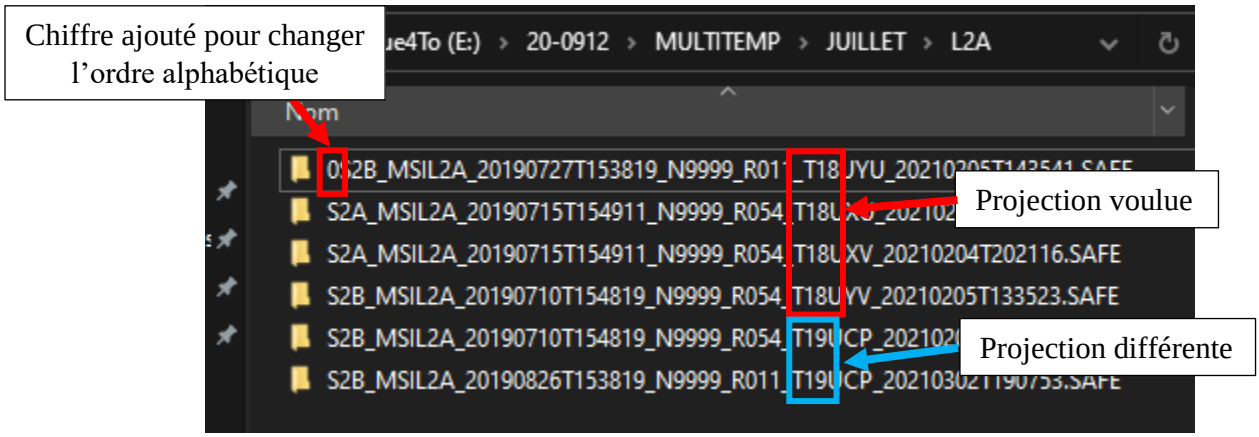
Les images étant importées séquentiellement d'après leur ordre alphabétique dans le dossier, il est nécessaire de placer en premier les images étant dans la projection finale voulue (modifier le nom du dossier .SAFE pour modifier l'ordre des images dans le dossier).

<u>Exemple :</u>

Images L2A brutes :



Images L2A dans le bon ordre :



2_Préparation des images : Nettoyage des nuages présents dans les images, reprojection et mosaïque.

Les chemins d'accès au .shp du territoire d'intérêt ainsi que des rasters correspondants doivent être renseignés.

Étape de Nettoyage des bandes : Si le territoire d'intérêt est entièrement couvert par une tuile Sentinel-2, mettre le paramètre « EXPORT » à « TRUE ».

Si plusieurs images sont nécessaires pour couvrir le territoire d'intérêt et que plusieurs projections sont présentes :

- Étape de reprojection : Définir la projection voulue (Par défaut la première image de juillet).
- Étape de mosaïquage : Mosaïque des différentes images et exporte chaque bande dans le dossier EXPORT de chaque mois.

3_Cartographie : Classification du territoire d'intérêt. Le paramètre à modifier est le nombre d'images à utiliser dans la classification en fonction du modèle choisi (« 2 » ou « 5 » images). La classification sera exportée dans le dossier « CLASSIFICATION » sous format raster.