
Cartographie des surfaces pastorales à l'aide des données Sentinel 2 L3A et des données ouvertes

Promesses et réalités

**Urcel Kalenga Tshingomba^{1,2}, Magali Jouven², Lucile Sautot¹,
Imad Shaqura², Maguelonne Teisseire¹**

1. TETIS, Univ Montpellier, AgroParisTech, CIRAD, CNRS, INRAE, Montpellier,
500, rue Jean-François Breton, 34093 Montpellier Cedex 5, France
urcel.kalengatshingomba@agroparistech.fr ; lucile.sautot@agroparistech.fr ;
maguelonne.teisseire@inrae.fr

2. SELMET, Univ Montpellier, Institut Agro / Montpellier SupAgro, CIRAD,
INRAE, Montpellier, France
2 Place Pierre Viala, 34060 Montpellier
Imad.Shaqura@inrae.fr ; magali.jouven@supagro.fr

RÉSUMÉ. Dans cet article, les auteurs expérimentent une démarche permettant de produire une cartographie cohérente de l'occupation des sols des surfaces des parcours en zones périméditerranéennes françaises représentées par les régions Occitanie et Provence-Alpes-Côte d'Azur. Quatre différentes sources de données sont utilisées : l'occupation des sols millésime OSO (OSO), le Registre parcellaire graphique (RPG), la BD-Forêt V.2.0 et les données satellites Sentinel 2 L3A. Le RPG de 2019 et la BD-Forêt actualisée en 2018 ont été utilisés comme principale source de données de référence pour l'entraînement des modèles en vue de classer les objets OSO 2019 de faible F-score, après extraction des variables spectrales, et des indices spectraux et texturaux issus des données Sentinel 2 L3A. Trois différentes tailles de données de référence ont été constituées. Ensuite 6 modèles ont été entraînés en utilisant l'algorithme Random Forest (RF) dont 3 modèles à partir des 3 jeux de données, intégrant toutes les variables extraites (98) et 3 modèles en reprenant uniquement les variables importantes (30) définies par le RF. Les résultats montrent des précisions globales stables pour tous les jeux de données utilisés et produisent une meilleure discrimination de 3 classes sur 4 : les pelouses, les cultures pérennes et les forêts. Les landes ne sont pas bien discriminées à cause de leur forte hétérogénéité spatiale.

ABSTRACT. In this paper we experiment an approach allowing to produce a coherent land use and cover of rangelands in the French Mediterranean areas represented by the Occitania and Provence-Alpes-Côte d'Azur regions. Four different data sources are used: the OSO land use and cover (OSO), the Registre Parcellaire Graphique (RPG), the BD-Forêt V.2.0 and Sentinel 2 L3A satellite data. The RPG of 2019 and the BD-Forêt updated in 2018 were used as the main source of reference data for training the models to classify the OSO 2019 low F-score class objects, after

extraction of the reflectances, spectral and textural indices variables derived from Sentinel 2 L3A data. Three different sizes of reference data are constructed. Then six models were trained using the Random Forest (RF) algorithm: 3 models from the 3 datasets integrating all extracted variables (98) and 3 models considering only the important variables (30) defined by the RF. The results show stable overall accuracies for all datasets used and produce better discrimination of three out of four classes: grasslands, perennial crops and forests. Moorlands are not well discriminated due to their high spatial heterogeneity.

MOTS-CLÉS : Random Forest, OSO 2019, RPG 2019, BD-Forêt, parcours.

KEYWORDS: Random Forest, OSO 2019, RPG 2019, Rangelands.

DOI: 10.3166/fig.2021.00112 © 2021 Lavoisier

1. Introduction

Dans les régions méditerranéennes européennes, les parcours occupent une grande part des surfaces pâturées (Bornard et Corzic, 2000 ; Bounejmate *et al.*, 2004). Ils représentent les surfaces de végétation non soumises aux pratiques culturales et dont les ressources, spontanées, sont principalement dédiées au pâturage des animaux herbivores. Les parcours sont très variés, allant des pelouses (végétations majoritairement herbacées) aux bois (végétations comportant un recouvrement important d'arbres), en passant par les landes (végétations comportant un recouvrement important de ligneux bas), avec des degrés de fermeture du milieu variables. Ceux-ci impactent plus ou moins la ressource pastorale consommable par les troupeaux, selon les essences présentes, l'accessibilité des plantes ligneuses et leur répartition spatiale (Zarovali *et al.*, 2007 ; Jouven *et al.*, 2010 ; Lasanta *et al.*, 2019).

L'utilisation de ces parcours par les troupeaux d'herbivores répond à des enjeux socio-économiques, écologiques et techniques variés. Les parcours comportent des ressources alimentaires variées (herbe, feuillages, fruits) qui fournissent une alimentation diversifiée et à moindre coût aux troupeaux, avec une flexibilité d'utilisation (Rancourt *et al.*, 2006 ; Jouven *et al.*, 2011). En contrepartie, l'élevage pastoral réalisé sur ces parcours permet de favoriser la biodiversité animale et végétale, maintenir les milieux ouverts et lutter contre les incendies, et préserver les propriétés des sols (Perevolotsky et Seligman, 1998 ; Akasbi *et al.*, 2012 ; Riedel *et al.*, 2013 ; UICN, 2014). Sous réserve d'une conduite raisonnée et adaptative du pâturage, le pastoralisme représente un moyen efficace permettant de valoriser ces milieux, et par conséquent, d'en tirer profit pour l'obtention de produits d'origine animale de qualité (Bornard et Corzic, 2000 ; Colditz, 2015).

La diversité des parcours, combinée parfois à un usage irrégulier par les troupeaux, rend difficile la caractérisation de ces milieux. Or pour raisonner la conduite du pâturage sur ces surfaces, une connaissance assez précise du type de végétation, en particulier de la présence et de la répartition spatiale de ligneux, est nécessaire. En fonction de la densité et de la répartition des ligneux, les ressources pastorales des landes peuvent être consommées différemment selon les espèces qui les pâturent (Benavides *et al.*, 2009 ;

Jáuregui *et al.*, 2009 ; Ferreira *et al.*, 2013). Au niveau de l'exploitation, l'élevage pastoral est souvent mobile et susceptible de varier les espaces utilisés d'une année à l'autre suivant la répartition spatiale des ressources. Une caractérisation des parcours à cette échelle est nécessaire afin de fournir une base de connaissance de leur variabilité spatio-temporelle, en vue de faciliter leur gestion à travers l'élevage pastoral. Une caractérisation des parcours au niveau territorial permettra également d'envisager des stratégies de gestion de l'espace pour répondre aux enjeux écologiques et préserver les services écosystémiques.

Les avancées dans le domaine de la télédétection offrent aujourd'hui des possibilités de caractériser la végétation à de larges échelles avec plus de précision qu'il y a 5 ans. Pour des études sur de grandes étendues géographiques, les données issues des satellites Sentinel 2¹ (lancés depuis 2015) possèdent de bonnes résolutions spatiale, spectrale (4 bandes de 10 m, 6 bandes de 20 m et 3 bandes de 60 m) et temporelle (passage tous les 5 jours). Ces caractéristiques permettent de suivre la végétation avec plus de détails que les données issues d'autres satellites tels que Landsat (8 bandes de 30 m de résolution spatiale, 1 bande panchromatique de 15 m et 2 bandes de 100 m et cycle de passage de 16 jours). La combinaison des caractéristiques des données Sentinel 2 est une opportunité pour caractériser la végétation des parcours en les croisant avec des sources de données de référence précises. L'occupation des sols OSO (avec 23 classes et 0,1 ha d'unité minimale de cartographie), produite à partir des séries temporelles d'images Sentinel 2 (Inglada *et al.*, 2017 ; Derksen *et al.*, 2020) constitue une donnée de meilleure précision au niveau national français. Elle est mieux que Corine Land Cover (European Environment Agency, 2007) qui malgré sa riche nomenclature intégrant à la fois l'occupation et l'usage des sols, n'est pas actualisée tous les ans et possède 25 ha d'unité minimale de cartographie. Cela limite ses applications dans plusieurs domaines, notamment pour la gestion du pâturage. Cependant, dans l'occupation des sols OSO, les classes pastorales (landes et pelouses) sont identifiées avec moins de précision que la plupart des classes (F-scores respectifs de 0,59 et 0,66). Cela peut également constituer une limite pour des applications pastorales qui utilisent souvent ces types de classes.

Pour mieux cartographier la végétation des parcours de la communauté de commune du Haut Verdon-Annot Saint-André (611 km²), Shaqura et Lasseur (2019) ont réalisé une classification bayésienne par objet à partir des images à très haute résolution spatiale (THRS) acquises par SPOT 6. Ils ont obtenu un coefficient Kappa de 0,84. Dans un système pastoral au Nord-Est du Portugal (660 km²), Castro *et al.* (2020) ont pu également cartographier 8 occupations des sols dont 3 occupations de parcours (Kappa = 0,89). Ils ont quant à eux traité les données de haute résolution spatiale (HRS) de Sentinel 2 L2A selon une approche par pixel, à l'aide de l'algorithme Random Forest (RF). La disponibilité des données Sentinel 2 (Drusch *et al.*, 2012), pour couvrir de grandes étendues, avec une grande fréquence d'acquisition (5 jours) (Whitcraft *et al.*, 2015), du RPG et de la BD-Forêt comme source de données de référence, est une opportunité pour le développement d'outils de traitement adaptés pour la caractérisation et le suivi des parcours. En vue de garantir la cohérence avec les données OSO, de nouvelles méthodes de traitement peuvent

1. <https://sentinel2.cnes.fr/fr>

y être appliquées dans le but de réduire les confusions qui existent entre les classes, notamment entre les classes de parcours, les cultures pérennes et les forêts.

Dans ce contexte, notre étude a pour objectif d'expérimenter une méthode permettant d'améliorer les précisions des classes pastorales de l'occupation des sols OSO à l'échelle des régions Occitanie et Provence-Alpes-Côte d'Azur (PACA). Pour cela, nous avons utilisé 4 sources de données : les images satellites Sentinel 2 de juillet 2019 comme variables prédictives, le RPG de 2019 et la BD-Forêt actualisée en 2018 comme données de référence, et l'occupation des sols OSO de 2019 dont nous visons d'améliorer les précisions des classes pastorales. Nous avons défini les méthodes d'analyse pour la production des modèles basés sur l'algorithme Random Forest (RF), que nous avons comparés en vue de rechercher les meilleurs performances pour la caractérisation des différents types de parcours.

2. Matériel et méthode

2.1. Milieu d'étude

Le milieu d'étude considéré pour ce travail représente environ 105 000 km² et concerne les deux régions Occitanie et Provence-Alpes-Côte d'Azur, situées dans la région méditerranéenne française selon les critères européens liés à la structure de la végétation (Quezel, 1999) (figure 1).

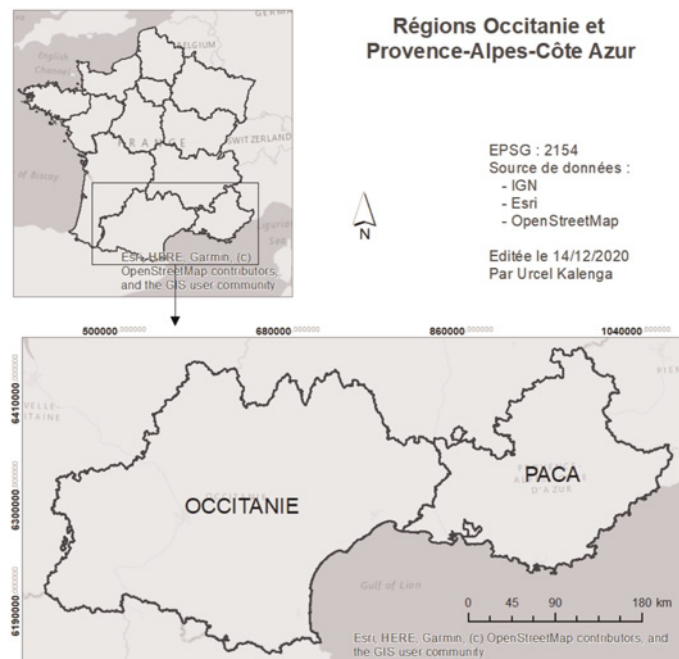


Figure 1. Position géographique de la zone d'étude

Le climat dans ce milieu est caractérisé par des hivers doux et des étés chauds, un ensoleillement important et des vents violents fréquents. Les pluies sont irrégulièrement réparties sur peu de jours. Les hivers et les étés sont secs et les printemps et les automnes très arrosés, souvent sous forme d'orages. On observe généralement en automne des épisodes méditerranéens occasionnant des précipitations intenses (plus de 200 mm en 24 heures). Ces situations météorologiques génératrices de fortes pluies sont de deux types : celles où l'influence du relief est prépondérante, observées dans des lieux tels que les Cévennes (épisodes Cévenols), les Pyrénées, les Alpes ou la Corse ; et celles qui sont exclusivement liées aux conditions météorologiques, observées dans des lieux tels que le Gard, l'Ardèche, l'Hérault ou la Lozère².

Les conditions météorologiques dans la région d'étude, en particulier la présence des nuages et des pluies, influencent la qualité des données satellites. Au printemps, en automne et en hiver, les conditions météorologiques empêchent souvent d'obtenir des scènes d'images totalement exemptes de nuages pour couvrir de larges étendues. Par contre, l'absence de couverture nuageuse et l'ensoleillement élevé durant l'été, permettent de meilleures conditions pour produire des images de qualité.

2.2. Données

Pour réaliser cette étude, nous avons utilisé 4 sources de données : les données satellites de synthèse mensuelle de Sentinel 2 L3A³ de juillet 2019, l'occupation des sols OSO de 2019 (OSO 2019)⁴, le Registre parcellaire graphique de 2019⁵ (RPG 2019), et la BD-Forêt version 2.0⁶ actualisée fin 2018.

L'occupation des sols OSO 2019 est une donnée de meilleure précision spatiale produite à grande échelle (nationale française) comparée à Corine Land Cover (European Environment Agency, 2007). Elle est actualisée tous les ans depuis 2016 par le Centre national d'études spatiales (CNES) en collaboration avec des laboratoires de recherche spécialisés⁷. Ce référentiel, produit à partir des séries temporelles des données Sentinel 2 (Inglada *et al.*, 2017), est une donnée de qualité homogène à l'échelle nationale de la France. Cependant, les classes de parcours ne sont pas bien discriminées à cause de leur forte hétérogénéité spatiale. Il s'agit des landes ligneuses et des pelouses qui possèdent des F-scores respectifs de 0,59 et 0,66 (*cf.* figure 3). Nous définissons le F-score au point 2.3.2 où nous présentons les critères d'évaluation des performances des modèles de classification. Dans cette étude, nous expérimentons les méthodes d'analyse pouvant permettre d'améliorer les précisions de ces classes de parcours.

2. www.meteofrance.fr

3. <https://labo.obs-mip.fr/multitemp/theias-sentinel-2-l3a-monthly-cloud-free-syntheses/>

4. <https://www.theia-land.fr/la-carte-doccupation-des-sols-millesime-2019-france-entiere/>

5. <https://geoservices.ign.fr/documentation/donnees/vecteur/rpg>

6. <https://geoservices.ign.fr/documentation/donnees/vecteur/bdforet>

7. <https://www.theia-land.fr/ceslist/ces-occupation-des-sols/>

La donnée OSO 2019 offre deux avantages intéressants pour cette étude. Premièrement, elle fournit de meilleures précisions pour la plupart des classes d'occupation des sols et des groupes de classes. Cela offre une possibilité de définir une approche de classification basée uniquement sur les classes de faible précision. Deuxièmement, la disponibilité de cette donnée en mode vectoriel permet de réaliser les analyses et les traitements en mode objet. Sans cela, il aurait été nécessaire d'envisager des méthodes de segmentation d'image, beaucoup plus coûteuses en temps de calcul.

Le RPG 2019 et la BD-Forêt 2.0 ont été utilisés comme principales sources de données d'entraînement et de test pour classifier les objets OSO 2019 correspondant aux classes de parcours. Le RPG est un système d'information géographique permettant l'identification des parcelles agricoles, géré par l'Agence de service et de paiement (ASP). Il fournit des informations détaillées sur l'occupation des sols et les structures foncières. Plusieurs classes de parcours (pelouses, landes ouvertes, landes fermées et forêts pâturées) y sont représentées, associées aux surfaces déclarées par les éleveurs dans le cadre de la Politique agricole commune (PAC). La BD-Forêt constitue quant à elle le référentiel géographique de description des essences forestières. Elle décrit les formations végétales forestières et naturelles par une approche de la couverture du sol traduisant une description de la densité de couvert du peuplement, de sa composition et de l'essence dominante. Elle possède une unité cartographique minimale de 5 000 m² (soit 0,5 hectare).

2.2.1. Données Sentinel 2

La période idéale pour effectuer le choix des données Sentinel 2 s'étend de mi-juin à mi-juillet. Ce choix tient à la fois sur la faible proportion de couverture nuageuse mensuelle et la possibilité de discriminer les classes pastorales. Le choix de cette période d'acquisition des images a été également réalisé par Castro *et al.* (2020).

La couverture de l'ensemble de la zone d'étude a requis 21 tuiles de données Sentinel 2. Premièrement, nous avons collecté les données Sentinel 2 de réflectance de surface (niveau L2A), résultat d'une correction atmosphérique des données Sentinel 2 non corrigées (niveau L1C), à partir de la chaîne de traitement MAJA développée par le Centre d'Etudes Spatiales de la Biosphère (CESBIO), le Centre National d'Etudes Spatiales (CNES) et l'Agence Spatiale Allemande (DLR)⁸. La collecte de ces données a été réalisée pour la période du 01 juin au 15 juillet 2019. Malgré la période d'acquisition d'image définie, il n'a pas été possible d'obtenir des tuiles d'images totalement dépourvues de couvertures nuageuses. Il a donc été nécessaire de réaliser les traitements sur les données Sentinel 2 de niveau L3A, sans couverture nuageuse, produites par la chaîne de traitement WASP. Ce sont des données de synthèse mensuelle produites en choisissant parmi plusieurs images de niveau L2A sur une période de 45 jours, le meilleur pixel dépourvu de nuage. Vingt-et-une tuiles des données Sentinel 2 L3A de juillet 2019 ont ainsi été collectées.

8. <https://labo.obs-mip.fr/multitemp/theias-sentinel-2-l3a-monthly-cloud-free-syntheses/>

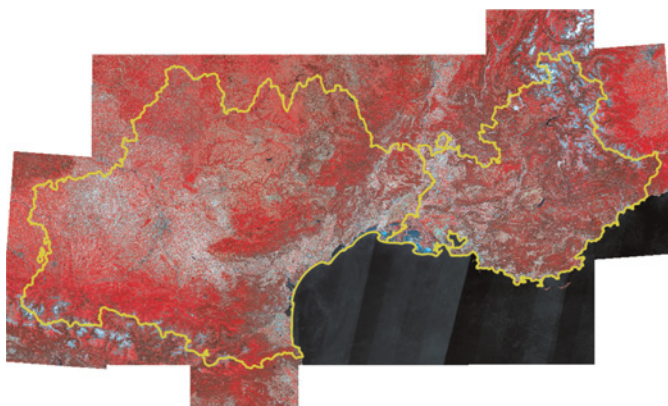


Figure 2. Composition colorée infrarouge fausses couleurs de la mosaïque de 21 tuiles de Sentinel 2 L3A couvrant les régions Occitanie et PACA.
Source de données : www.theia-land.fr

Les données Sentinel L3A ne représentent pas fidèlement l'information temporelle car il est difficile de savoir la date exacte pour chaque pixel. Elles sont principalement utilisées pour la visualisation. Dans l'état actuel, la plupart des travaux se réfèrent plutôt à l'utilisation des données Sentinel 2 L2A. Cependant, dans cette étude, nous considérons que sur la période de 45 jours (nécessaires pour produire les données Sentinel 2 L3A), la variabilité des propriétés spectrales des classes n'impacte pas significativement leur discrimination.

Les prétraitements réalisés sur les données Sentinel 2 L3A (cf. figure 4B) ont concerné le rééchantillonnage des bandes spectrales de 20 m de résolution spatiale à 10 m de résolution spatiale et l'empilement des 10 bandes spectrales pour toutes les 21 tuiles. Ensuite, un mosaïquage a été réalisé pour obtenir une couverture totale de la zone d'étude (cf. figure 2).

2.2.2. Occupation des sols Millésime OSO de 2019

Les données OSO de 2019, constituées de 23 classes (cf. figure 3), ont été obtenues sous forme vectorielle afin de faciliter les traitements. L'unité minimale de cartographie des données OSO est de 0,1 ha, obtenue par régularisation, vectorisation puis simplification de la classification en mode raster de 10 m de résolution spatiale (Inglada *et al.*, 2017). Parmi les 23 classes, celles qui sont susceptibles d'être pâturées sont les pelouses, les landes ligneuses, les prairies, les vergers, les vignes et les forêts. Le regroupement des classes en groupes⁹ : urbain (bâtis dense, bâtis diffus, zones industrielles et commerciales et surfaces routières), cultures annuelles (colza, céréales, protéagineux, soja, tournesol, maïs, riz et tubercules), cultures pérennes (vergers et vignes), forêts (de feuillus et de conifères), et formations naturelles basses (pelouses et landes ligneuses) (cf. figure 3) ; pour l'occupation des sols de 2017 a permis de réduire

9. <http://osr-cesbio.ups-tlse.fr/oso/>

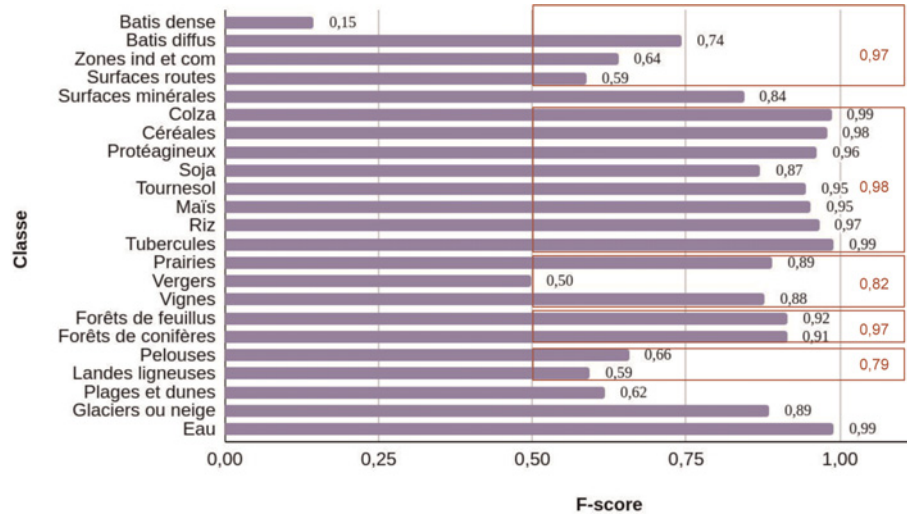


Figure 3. F-scores des 23 classes de l'occupation des sols OSO de 2019 et des classes regroupées (en rouge) de l'occupation des sols OSO de 2017. Les F-scores des classes regroupées d'OSO 2017 ont été utilisés à défaut de la disponibilité de l'information pour la donnée 2019. Source : [http:// osr-cesbio.ups-tlse.fr/~oso/](http://osr-cesbio.ups-tlse.fr/~oso/)

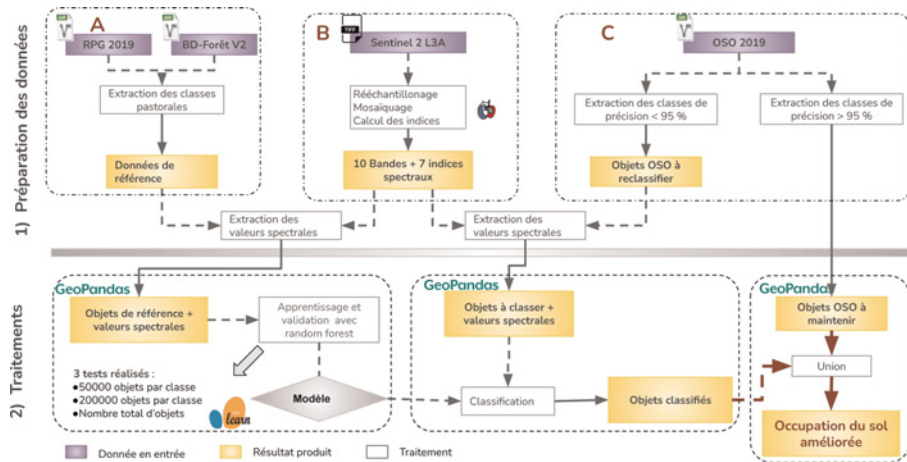


Figure 4. Schéma général de traitement d'amélioration de la précision de la classification des classes pastorales et des cultures pérennes. Avec comme données en entrée : le RPG 2019 et la BD-Forêt (4A), l'occupation des sols OSO 2019 (4C) et les données Sentinel 2 L3A (4B)

la confusion de la classification. Cela améliore les F-scores des zones artificialisées (urbain, 0,97), des cultures annuelles (0,98) et des forêts (0,97). Les groupes « cultures pérennes » et « formations naturelles basses » possèdent des F-scores qui nécessitent d'être améliorés (respectivement 0,82 et 0,79). Les classes « Surfaces minérales », « Glaciers ou neige » et « Eau » possèdent de grandes valeurs de F-score individuellement (respectivement 0,84 ; 0,89 et 0,99). La classe « Plages et Dunes » possède quant à elle un faible F-score (0,62).

La lecture de la matrice de confusion d'OSO nous a permis d'effectuer le choix des objets à reclasser. Étant donné l'indisponibilité (dans nos recherches effectuées) de la matrice de confusion pour OSO 2019, nous avons utilisé la référence de 2018. Dans la matrice de confusion d'OSO 2018 (cf. tableau 1), les landes ligneuses et les pelouses sont en confusion avec les vergers et les vignes et beaucoup plus avec les forêts. La grande partie d'objets qui devraient appartenir à ces deux classes est attribuée aux autres classes (faux négatifs) et elles possèdent également une part importante d'objets appartenant aux autres classes (faux positifs). Par ailleurs, même si les vergers possèdent un faible taux d'objets appartenant à d'autres classes, une part importante de leurs objets se trouvent repartis dans toutes les autres classes, principalement les forêts de feuillus. Ainsi, afin de réduire les imprécisions des pelouses, landes ligneuses et vergers dans OSO, il est nécessaire d'inclure les autres classes avec lesquelles elles sont le plus en interaction, dont les vignes et les deux types de forêt. Cependant, d'autres parts non négligeables d'imprécisions, bien que minimes, sont dues à d'autres classes qui n'ont pas été retenues pour la classification. Pour ces raisons, il est impossible d'améliorer intégralement les faux négatifs et les faux positifs de ces classes. Seuls les faux négatifs et les faux positifs liés aux classes maintenues pourraient être améliorés.

Dans cette étude, nous avons scindé la donnée OSO 2019 en deux (cf. figure 4C) :

1) les classes ou groupes de classes dont le F-score est supérieur ou égal à 0,95 et n'ayant pas trop d'influence sur les classes de facile précision, à maintenir, et 2) les groupes de classes de faibles F-scores destinés à être améliorés par des traitements spécifiques en même temps que les autres classes bien discriminées avec lesquelles elles sont le plus confondues. Les classes que nous retenons pour l'amélioration sont celles des pelouses (F-score = 0,66), des landes ligneuses (F-score = 0,59), des vergers (F-score = 0,50), des vignes (F-score = 0,88), des forêts de feuillus (F-score = 0,92) et de conifères (F-score = 0,91). La classe « plages et dunes » (F-score = 0,62) quant à elle, n'a pas été retenue pour être classifiée à cause du manque de données de référence. Toutefois, le fait de ne pas intégrer cette classe dans le processus de classification n'affecte pas les précisions des classes de végétation retenues car étant spectralement séparées.

Pour cette étude, nous avons tenté de distinguer les surfaces de parcours en fonction de la densité des ligneux. Ainsi, nous avons défini deux classes de landes ligneuses : les landes ouvertes et les landes fermées. Cette catégorisation des landes a été motivée par l'existence dans les données de référence du RGP 2019 des objets caractérisant les landes en fonction de leur densité de ligneux (tableau 2).

Au total, 1 888 034 objets à classer ont été extraits de l'occupation des sols OSO 2019.

Tableau 1. Extrait de la matrice de confusion d'OSO 2018 sur les classes à faibles F-scores et celles qui influencent le plus la qualité de leurs précisions.

Classe	Ver.	Vig.	F. F	F. C.	Pel.	L. L	Rap.	F-S.
Ver.	2,53E + 05	4,23E + 04	7,11E + 04	1,12E + 04	3,57E + 04	1,89E + 04	0,38	0,53
Vig.	3,18E + 03	2,73E + 06	6,00E + 03	3,88E + 04	4,22E + 04	2,55E + 04	0,87	0,87
F. F	-	-	3,42E + 07	1,14E + 06	4,23E + 05	3,60E + 05	0,94	0,91
F. C.	-	-	1,41E + 06	3,44E + 07	6,17E + 05	3,80E + 05	0,92	0,91
Pel.	-	6,07E + 04	1,04E + 06	1,01E + 06	1,24E + 07	2,25E + 06	0,64	0,65
L. L	-	4,39E + 04	1,08E + 06	8,63E + 05	3,09E + 06	7,36E + 06	0,52	0,57
Pré.	0,84	0,88	0,89	0,9	0,66	0,64		

Ver. = verger, Vig. = vigne, F. F. = forêt de feuillus, F.C. = forêt de conifères, Pel. = pelouse, L. L. = landes ligneuses, Pré. = précision, Rap. = rappel, et F.-S. = F-score

Tableau 2. Correspondance entre les classes du RPG 2019 et les classes pastorales (hormis les bois pâturés) maintenues pour l'apprentissage des modèles de classification

Code culture	Libellé culture	Classe correspondante
BOP	Bois pâturé	Bois pâturé
PPH	Prairie permanente herbe prédominante (ressources fourragères ligneuses absentes ou peu présentes)	Pelouse et prairie
PRL	Prairie en rotation longue (6 ans ou plus)	Pelouse et prairie
SPH	Surface pastorale herbe prédominante et ressources fourragères ligneuses présentes	Lande ouverte
SPL	Surface pastorale ressources fourragères ligneuses prédominantes	Lande fermée

2.2.3. Registre parcellaire graphique de 2019 et BD-Forêt version 2.0

Dans le RPG 2019, les informations correspondant aux classes OSO 2019 à cartographier sont celles qui concernent les pelouses, les landes ouvertes, les landes fermées et les cultures annuelles qui y sont bien représentées en termes d'effectif d'objets. Le tableau 2 montre la correspondance entre le champ « Code Culture » du RPG 2019 et les classes de végétation de parcours correspondantes. La classe « Bois pâturé » n'a pas été utilisée dans le processus de classification car elle se réfère aux classes forestières avec lesquelles il ne sera pas possible de la distinguer spectralement. Ces types de forêts sont utilisés par les troupeaux pour pâturer le sous-bois et s'abriter. Cependant, les objets de cette classe pourront être intégrés à la fin du processus de traitement pour représenter les espaces forestiers utilisés par les troupeaux, vu leur importance pour le pastoralisme.

Tous les objets du RPG 2019 correspondant aux classes de faible F-score dans OSO 2019 ont été extraits pour constituer la base de données de référence et d'évaluation des précisions. Quatre classes ont été définies : « Pelouse et Prairie permanente », « Lande ouverte », « Lande fermée » et « Culture pérenne ». La classe « Pelouse et prairie permanente » a été appelée « Pelouse » pour simplifier la nomenclature. Au total, 890 826 objets de référence ont été extraits du RPG 2019. Ce nombre d'objets a ensuite été réduit à 887 854 après élimination des géométries invalides.

La BD-Forêt a été utilisée comme donnée de référence pour les forêts. Elle contient cinq types d'espaces naturels et semi-naturels : les forêts fermées, les forêts ouvertes, les peupleraies, les landes (maquis, garrigues), et les formations herbacées (pelouses alpines, pelouses pastorales). Une extraction des classes forestières (forêts fermées et

forêts ouvertes) a permis d'obtenir 413 816 objets de référence pour la classe forestière. Ce nombre d'objets est passé à 412 705 après élimination des géométries invalides.

2.3. Traitements

2.3.1. Variables prédictives

Les traitements réalisés dans cette étude (*cf.* figure 4) s'appliquent pour améliorer la discrimination des classes de végétation de parcours, des cultures pérennes et des forêts extraits d'OSO 2019 (*cf.* figure 4). Les objets correspondant à ces classes dans le RPG 2019 et la BD-Forêt (*cf.* figure 4A) sont utilisés comme référence. Trois groupes de variables ont été produits : les reflectances, les indices spectraux (*cf.* tableau 3) et les indices de texture. Les statistiques de géométrie ne permettent pas de caractériser les classes à cause de la grande hétérogénéité de la forme de leurs objets.

Pour les objets OSO et ceux de référence (RPG et BD-Forêt combinés), les moyennes et les écarts types des 10 bandes spectrales de Sentinel 2 L3A (20 variables), ceux de 4 indices de végétation et 3 indices de sol (14 variables) calculés, ainsi que ceux des 8 indices de texture d'Haralick (Haralick *et al.*, 1973) (moyenne, variance, homogénéité, contraste, dissimilarité, entropie, second moment, et corrélation) calculés pour les 4 bandes de 10 m de pixel (64 variables), ont été respectivement extraits. L'utilisation des reflectances et des indices spectraux a été motivé par le travail de Castro *et al.* (2020), qui ont pu mieux caractériser les classes de parcours à partir de ces variables. Ce processus a permis d'obtenir 98 variables prédictives pour l'entraînement et la validation des modèles, et pour la classification des objets OSO 2019. Tous les calculs d'indices spectraux ont été réalisés sous Python associé à la boîte à outils Orfeo Toolbox 7.2 (OTB 7.2)¹⁰. Les calculs de texture ont été réalisés avec logiciel ENVI 5.6.1¹¹.

2.3.2. Définition des classes

Bien avant de procéder au traitement de classification supervisée, il a été nécessaire d'analyser la séparabilité des cinq classes retenues en tenant compte des différents groupes de variables sélectionnées. Nous avons appliqué l'analyse discriminante linéaire (ADL) implémentée dans la librairie Scikit-Learn¹² pour analyser la séparabilité des classes en fonction des trois groupes de variables (les reflectances (20), les indices spectraux (14) et les indices texturaux (64)). L'ADL permet d'effectuer une réduction supervisée de la dimensionnalité, en projetant les données d'entrée dans un sous-espace linéaire constitué des directions qui maximisent la séparation entre les classes. La plupart des études utilisent l'ADL pour des fins de classification. Cependant, elle est également utile pour effectuer des analyses préliminaires de séparabilité de classe. Nous

10. <https://www.orfeo-toolbox.org/>

11. <https://www.l3harrisgeospatial.com/Software-Technology/ENVI>

12. https://scikit-learn.org/stable/modules/generated/sklearn.discriminant_analysis.LinearDiscriminantAnalysis.html

Tableau 3. Bandes spectrales et indices de végétation et de sol dérivés de Sentinel 2 L3A, pris comme variables pour l'entraînement du modèle

Bandes et indices spectraux	Definition	Formule
B ₂	Bleue (490 nm)	
B ₃	Verte (560 nm)	
B ₄	Rouge (665 nm)	
B ₅	Red edge (705 nm)	
B ₆	Red edge (749 nm)	
B ₇	Red edge (783 nm)	
B ₈	Proche infrarouge (842 nm)	
B _{8a}	Proche infrarouge (865 nm)	
B ₁₁	Moyen infrarouge (1610 nm)	
B ₁₂	Moyen infrarouge (2190 nm)	
NDVI	Indice différentiel de végétation normalisé (Rouse <i>et al.</i> , 1974)	$\frac{B_8 - B_4}{B_8 + B_4}$
NDI ₄₅	Indice différentiel normalisé (Delegido <i>et al.</i> , 2013)	$\frac{B_5 - B_4}{B_5 + B_4}$
IRECI	Indice de chlorophylle inverse sur le Red-edge (Delegido <i>et al.</i> , 2013)	$\frac{B_7 - B_4}{B_5 / B_6}$
TNDVI	NDVI transformé (Senseman <i>et al.</i> , 1996)	$\left(\frac{B_8 - B_4}{B_8 + B_4} + 0,5\right)^{0,5}$
CI	Indice de coloration (Pouget <i>et al.</i> , 1990)	$\frac{B_4 - B_3}{B_4 + B_3}$
BI	Indice de brillance (Escadafal, 1989)	$\left(\frac{B_4^2 - B_3^2}{4}\right)^{0,5}$
RI	Indice de rougeur (Pouget <i>et al.</i> , 1990)	$\frac{B_4^2}{B_3}$

avons appliqué l'ADL sur les données de référence de trois départements (Alpes-de-Haute-Provence (04), Aude (11) et Var (83), en vue d'analyser les possibilités de maintenir ou de regrouper certaines classes ainsi que les variables sélectionnées.

Pour des raisons d'espace, nous présentons dans cette section uniquement les résultats du département 04 (figure 5). Les résultats des départements 11 (*cf.* figure 8) et 83 (figure 9) sont repris en annexe de ce travail. On observe dans les boîtes à moustaches

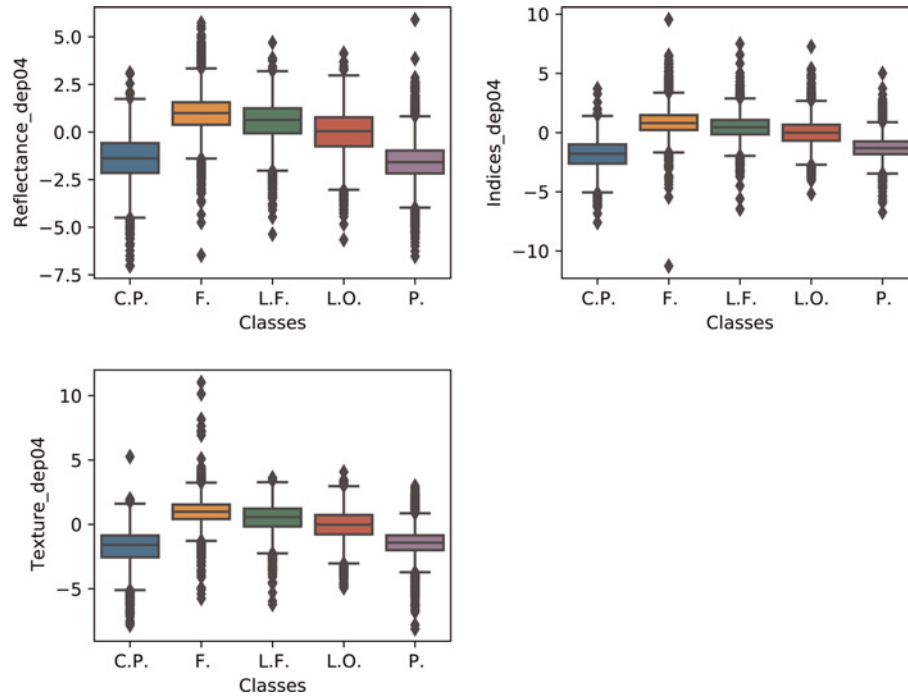


Figure 5. Boîtes à moustaches des premiers composants de l'ADL appliquée pour analyser la séparabilité des 5 classes retenues à partir des reflectances, des indices spectraux et texturaux, pour le département des Alpes-de-Haute-Provence (04)

des premiers composants de l'ADL que les classes retenues sont difficilement séparables quel que soit le groupe de variables. Les cultures pérennes sont plus confondues avec les pelouses, et les landes fermées sont en confusion entre les forêts et les landes ouvertes. Dans les figures 8 et 9 (en annexe), les cultures pérennes sont bien séparées des pelouses et des autres classes. Cela peut être justifié par la variabilité spatiales des classes. Enfin, les boîtes à moustaches présentent les tendances similaires pour les variables sélectionnées, montrant leurs corrélations à ce point de vue. Cependant, l'utilisation d'un algorithme performant pourrait tout de même permettre de tirer profit de ces informations pour maximiser les possibilités de discrimination.

À ce stade, nous estimons que les landes fermées, vu leur forte confusion avec les landes ouvertes et les forêts, risquent d'augmenter les imprécisions lors de la classification. De ce fait, nous avons opté pour le regroupement en une classe des landes fermées et ouvertes. Nous avons donc maintenu au final quatre classes : pelouse, lande, culture pérenne et forêt.

2.3.3. Sélection des jeux d'entraînement

Dans (Millard et Richardson, 2015), les auteurs montrent qu'un échantillonnage des données d'entraînement est bon s'il prend en compte la proportionnalité des différentes classes dans le milieu. Il est difficile de connaître les proportions des classes pastorales du RPG pour établir ce type d'échantillonnage. Dans les données de référence, le nombre d'objets correspondant à chaque classe est très varié. Les landes sont les plus minoritaires. Pour évaluer l'effet du poids des données d'entraînement pour chaque classe, nous avons constitué 3 tailles différentes des jeux de données d'entraînement. Dans un premier temps, les données de référence ont été scindées en deux : deux tiers pour l'entraînement et un tiers pour l'évaluation des modèles de classification. Les données d'entraînement ont par la suite été scindées en 3 jeux : 1) 50 000 objets pour chaque classe pour le premier, 2) 200 000 objets pour les classes majoritaires et la totalité d'objets pour les classes minoritaires pour le deuxième, et 3) l'ensemble d'objets de chaque classe pour le troisième. Les résultats de ce processus sont présentés dans le tableau 4.

2.3.4. Choix de l'algorithme de classification

Les données acquises sont de sources différentes. En plus, les données RPG sont produites par un grand nombre d'exploitants. Plusieurs facteurs d'imprécision de mesure peuvent subvenir pendant leur numérisation. Ces deux situations évoquées peuvent constituer une source de variabilité qui nécessite l'utilisation d'un algorithme performant. L'algorithme RF (Breiman, 2001) est connu pour sa robustesse face aux données non balancées (Khalilia *et al.*, 2011 ; Khoshgoftaar *et al.*, 2007) et provenant des sources différentes (Frenay et Verleysen, 2014). C'est un algorithme de classification et de régression qui réalise un apprentissage basé sur un ensemble d'arbres de décision à partir d'une sélection aléatoire des échantillons au niveau de chaque arbre (Breiman, 2001 ; Fawagreh *et al.*, 2014). Un tirage aléatoire avec remise des individus (*bootstrap*) est réalisé pour l'entraînement des arbres et les individus restants non tirés sont utilisés en interne en vue d'estimer la performance du modèle (Breiman, 2001 ; Belgiu et Dragut, 2016). Le RF est également stable quant au choix des paramètres de traitement (Pelletier *et al.*, 2016). Dans Lin *et al.* (2017), les auteurs ont également montré que le RF permet de mieux prédire les données massives, comparé à d'autres algorithmes de classification tel que le SVM. L'ensemble de ces avantages du RF nous ont motivé à l'utiliser pour entraîner les modèles en vue de discriminer les classes retenues.

2.3.5. Sélection des variables

La sélection des variables prédictives a été réalisée pour maximiser les possibilités de discriminer les classes sélectionnées. Ne sachant pas à l'avance les variables déterminantes pour produire les meilleures performances des modèles sur les 98 sélectionnées, une méthode appropriée de réduction de dimensionnalité des variables a été nécessaire. Parmi les méthodes de réduction de dimensionnalité, celles qui sont

Tableau 4. Nombre d'objets extraits pour chaque classe dans le RPG 2019 et la BD-Forêt V2.0 et 3 jeux de données utilisés pour l'entraînement et l'évaluation de l'algorithme de classification

Classe	Effectif	Entraînement	Test	Jeu 1	Jeu 2	Jeu 3
Pelouse	456 991	305 683	151 308	50 000	200 000	305 683
Lande	170 107	114 156	55 951	50 000	114 156	114 156
Culture pérenne	260 756	174 617	86 139	50 000	174 617	174 617
Forêt	412 705	276 919	135 786	50 000	200 000	276 919
Total	1 300 559	871 375	429 184	200 000	688 773	871 375

intégrées dans les algorithmes de *machine learning* présentent un avantage. Elles prennent en compte l'interaction avec le modèle de classification pour définir la liste des variables importantes (Saeys *et al.*, 2007). Nous avons utilisé les résultats de l'autoévaluation des performances du RF pour déterminer les variables importantes (Archer et Kimes, 2008 ; Breiman, 2001 ; Dewi et Chen, 2019b). Au cours de l'évaluation interne du modèle, l'algorithme attribue un score à chaque variable qui détermine son niveau d'importance dans l'apprentissage du modèle. Ce score d'importance est basé sur la décroissance moyenne de l'impureté lorsqu'une variable est choisie au niveau d'un nœud. De cette manière, les variables ayant des scores relativement élevés peuvent être choisies pour entraîner le modèle. En raison du caractère stochastique du RF, nous avons reproduit 20 fois l'entraînement du modèle sur le premier jeu de données (avec 50 000 objets par classe) pour définir les variables importantes à partir des moyennes et des écarts types de leurs scores. La figure 6 présente le classement décroissant des variables en fonction de leurs scores d'importance. Nous avons sélectionné comme variables importantes celles dont les scores sont égaux ou supérieurs à 0,009. Cette limite inférieure a été choisie pour deux raisons : le faible score d'importance et le niveau à partir duquel l'écart entre deux scores successifs devient significatif. Ainsi, 30 variables importantes sur les 98 ont été retenues.

2.3.6. Entraînement et évaluation des modèles de classification

Pour évaluer les performances produites en utilisant l'ensemble des 98 variables de base et les 30 nouvelles variables réduites, nous avons entraîné l'algorithme sur les trois jeux de données constitués sans réduction de dimensionnalité et sur les trois jeux de données avec réduction de dimensionnalité des variables prédictives.

Afin d'obtenir les meilleures performances de l'algorithme à entraîner sur les jeux de données préparés, une combinaison des valeurs de deux paramètres importants du RF a été évaluée : le nombre d'arbres de décision à considérer (100, 250, 500, 750, 1000, 1250, 1500, 1750, 2000) et le nombre des variables à considérer à chaque nœud (4, 8, 12, racine carrée du nombre total). La meilleure valeur du nombre d'arbres a été de 500 pour les deux types de données. Et celles des variables au niveau des nœuds, de 10 et 5 respectivement pour les données à dimensionnalité non réduite et les données à dimensionnalité réduite. Nous avons par la suite réalisé une validation croisée à 10 blocs afin d'évaluer la stabilité des modèles pour l'entraînement des données utilisées.

Pour évaluer la performance des modèles de classification, nous avons utilisé la précision globale et le F-score comme indicateurs. Le F-score permet de réaliser un meilleur compromis entre la valeur de la précision et celle du rappel à partir de leur moyenne harmonique. Ces trois mesures sont présentées par les formules suivantes (Van Rijsbergen, 1979) :

$$Précision = \frac{VP}{VP + FP} \quad (1)$$

$$Rappel = \frac{VP}{VP + FN} \quad (2)$$

$$F - score = \frac{(\beta^2 + 1) \times Précision \times Rappel}{(\beta)^2 \times Précision + Rappel} \quad (3)$$

Où VP est le nombre des résultats positifs correctement classés comme positifs, FP le nombre des résultats négatifs incorrectement classés comme positifs, FN le nombre de résultats positifs incorrectement classés comme négatifs et β le paramètre qui contrôle la balance entre la précision et le rappel. Lorsque $\beta > 1$, le F-score devient plus orienté vers le rappel et si $\beta < 1$, il devient plus orienté vers la précision. Pour cette étude, la valeur choisie pour β est de 2, définie par défaut dans Scikit-Learn.

Une comparaison des précisions globales et des F-scores a été faite pour sélectionner le jeu de données permettant d'obtenir une meilleure discrimination des classes de parcours.

Tous les traitements décrits dans cette sous section ont été réalisés grâce aux modules de la librairie Scikit-learn sous Python 3.8.

3. Résultats

3.1. Performances des modèles

Les résultats de l'entraînement des modèles de classification sont présentés dans les tableaux 5, 6 et 7 ci-après.

Tableau 5. Matrice de confusion et qualité de la prédiction produites à partir du modèle utilisant le jeu de données possédant 50 000 objets par classe

	Classes	Classification				
		Dimensionnalité réduite : précision globale de 0,79 ±0,002				
		Pelouse	Lande	Culture P.	Forêt	Rappel
Référence	Pelouse	121458	20808	5553	3489	0,80
	Lande	9074	34822	1685	10370	0,62
	Culture P.	4104	3356	76896	1783	0,89
	Forêt	3832	23354	1415	107185	0,79
	Précision	0,88	0,42	0,90	0,87	
	F-score	0,84	0,50	0,90	0,83	
		Dimensionnalité non réduite : précision globale de 0,80 ±0,002				
		Pelouse	Lande	Culture P.	Forêt	Rappel
	Pelouse	122613	19531	6321	2843	0,81
	Lande	9004	34892	1725	10330	0,62
	Culture P.	4779	3791	75960	1609	0,88
	Forêt	3779	21175	1201	109631	0,81
	Précision	0,87	0,44	0,89	0,88	
	F-score	0,84	0,52	0,89	0,84	

Tableau 6. Matrice de confusion et qualité de la prédiction produites à partir du modèle utilisant le jeu de données test 2

	Classes	Classification				
		Dimensionnalité réduite : précision globale de 0,82 ±0,002				
		Pelouse	Lande	Culture P.	Forêt	Rappel
Référence	Pelouse	131720	10234	4705	4649	0,87
	Lande	13218	25357	1821	15555	0,45
	Culture P.	4525	1669	77814	2131	0,90
	Forêt	4922	11409	1406	118049	0,87
	Précision	0,85	0,52	0,91	0,84	
	F-score	0,86	0,48	0,91	0,85	
		Dimensionnalité non réduite : précision globale de 0,83 ±0,003				
		Pelouse	Lande	Culture P.	Forêt	Rappel
	Pelouse	133196	9185	5054	3873	0,88
	Lande	13338	25607	1832	15174	0,46
	Culture P.	5278	1906	76962	1993	0,89
	Forêt	4884	9789	1248	119865	0,88
	Précision	0,85	0,55	0,90	0,85	
	F-score	0,86	0,50	0,90	0,87	

Tableau 7. Matrice de confusion et qualité de la prédiction produites à partir du modèle utilisant le jeu de données test 2

	Classes	Classification				
		Dimensionnalité réduite : précision globale de 0,83 ±0,003				
		Pelouse	Lande	Culture P.	Forêt	Rappel
Référence	Pelouse	137506	5356	3275	5171	0,91
	Lande	17170	18406	1450	18925	0,33
	Culture P.	6183	1024	76350	2582	0,89
	Forêt	6289	6624	1046	121827	0,90
	Précision	0,82	0,59	0,93	0,82	
	F-score	0,86	0,42	0,91	0,86	
		Dimensionnalité non réduite : précision globale de 0,83 ±0,003				
		Pelouse	Lande	Culture P.	Forêt	Rappel
	Pelouse	138773	4716	3456	4363	0,92
	Lande	17343	18677	1453	18478	0,33
	Culture P.	7368	1179	75196	2396	0,87
	Forêt	6137	5465	948	123236	0,91
	Précision	0,81	0,62	0,93	0,83	
	F-score	0,86	0,43	0,90	0,87	

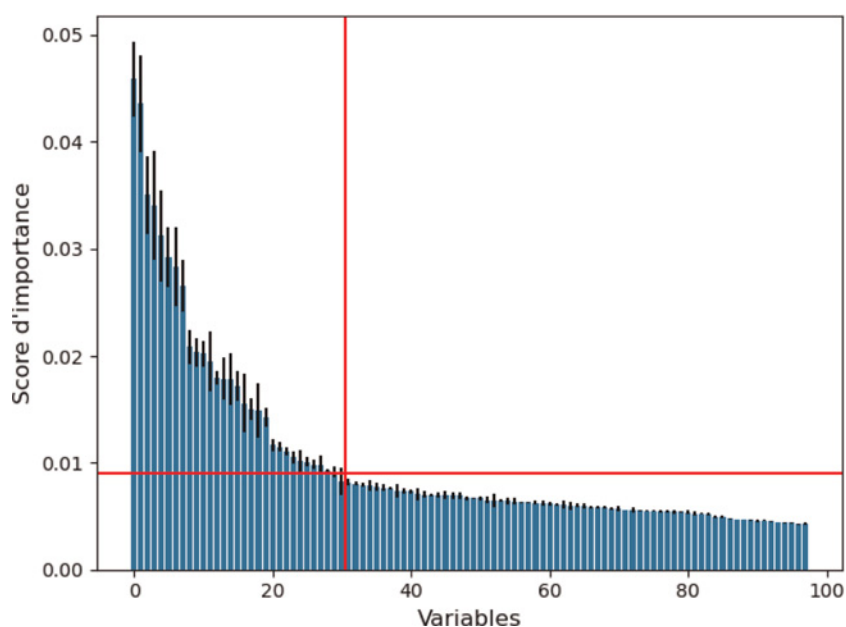


Figure 6. Classement des variables en fonction de leurs scores d'importance après 20 itérations sur le jeu de données avec 50 000 objets par classe

L'entraînement par le RF a produit des précisions globales stables pour les trois jeux de données avec et sans réduction des variables, suite à la validation croisée, leurs écarts types étant faibles. Les précisions globales produites pour les jeux de données avec réduction du nombre des variables sont respectivement $0,79 \pm 0,002$; $0,82 \pm 0,002$ et $0,83 \pm 0,003$. Celles des données sans réduction du nombre de variables sont respectivement $0,80 \pm 0,002$; $0,83 \pm 0,003$ et $0,83 \pm 0,003$. Les jeux de données à dimensionnalité non réduite ont produit de meilleures précisions globales comparé aux données à dimensionnalité réduite, sauf pour le jeu de donnée 3 (contenant l'ensemble d'objet des classes) où elles sont équivalentes.

Pour tous les jeux de données, les landes possèdent de faibles F-scores (de 0,42 à 0,50, et de 0,43 à 0,52 respectivement pour les jeux de données à dimensionnalité réduite et non réduite) comparées aux autres classes (pelouse, culture pérenne et forêt) qui possèdent des F-scores élevés (de 0,83 à 0,91, et de 0,84 à 0,90 respectivement pour les jeux de données à dimensionnalité réduite et non réduite). Les F-scores ont varié pour chaque classe selon les combinaisons de jeux de données d'entraînement et de variables. Aucune combinaison n'a été optimale, au sens d'une maximisation des F-scores pour toutes les classes. Il reste cependant difficile d'améliorer la discrimination des landes à cause de confusion avec toutes les autres classes.

En analysant les matrices de confusion, on observe une forte interaction entre les landes et les forêts, suivi des pelouses. Les cultures pérennes influencent moins la

discrimination des landes par rapport aux forêts et aux pelouses. Les F-scores de toutes les autres classes (sans les landes) sont élevés pour tous les jeux de données. Compte tenu des enjeux liés à la présence et à la répartition des ligneux dans les parcours, le choix du jeu de données et du nombre de variables serait orienté vers le modèle qui produit les meilleurs F-scores des landes. Il s'agit du jeu de donnée 1 (50 000 effectif d'objets d'entraînement pour chaque classe) sans réduction de dimensionnalité d'après les résultats obtenus. Il est néanmoins nécessaire de faire remarquer que la discrimination des landes n'a pas été meilleure autant que les forêts, les cultures pérennes et les pelouses. Le meilleur F-score des landes est de 0,52 dans nos résultats tandis que dans OSO 2019 il est de 0,57.

La réduction de dimensionnalité des variables a principalement affecté la classe des landes. Les F-scores de ces deux classes diminuent pour chaque jeu de données quand la dimensionnalité des variables est réduite. Les classes des pelouses et des cultures pérennes ne sont quant à elles pas impactées.

3.2. Occupation des sols produite

Le modèle ayant produit le meilleur F-score des landes a été utilisé pour reclasser les objets OSO extraits. Une visualisation des résultats sur deux différentes zones prises dans le département de l'Aude (11) est présentée à la figure 7. Nous faisons une

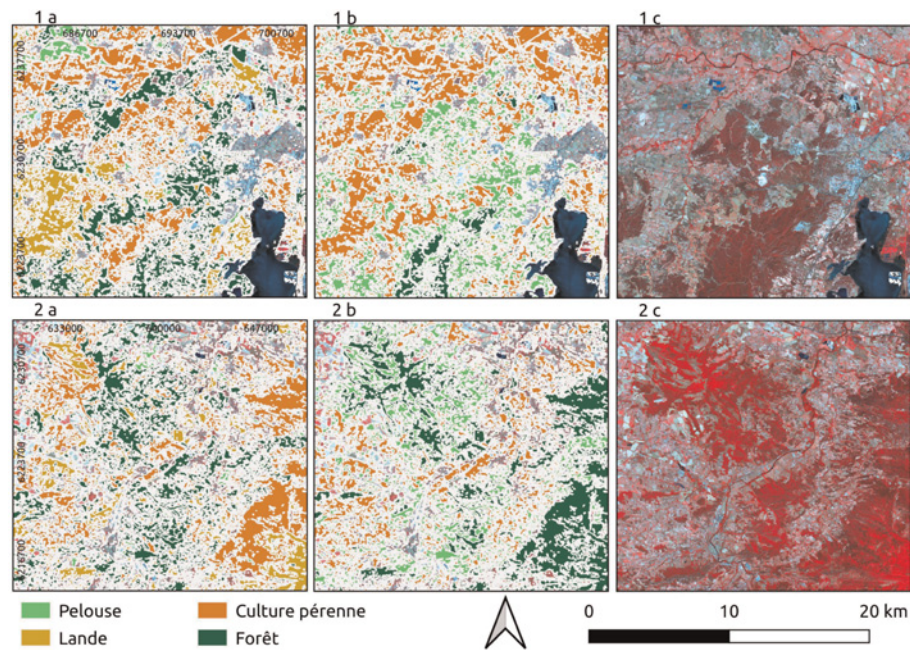


Figure 7. Comparaison entre l'occupation des sols produite, OSO 2019 et la composition colorée infrarouge fausses couleurs

comparaison de la qualité de la nouvelle classification avec l'occupation des sols OSO originale et l'image Sentinel 2 L3A.

Pour la première zone, le résultat produit (1a) montre une substitution des cultures pérennes vers le nord (1b) par les forêts. De même, les cultures pérennes (1b) ont été substituées par les landes dans la nouvelle classification (1a) à l'est de la zone. Les forêts qui apparaissent au centre et au sud de la zone ont substitué les pelouses. On observe néanmoins une correspondance pour les cultures pérennes entre les deux produits, notamment vers le nord et le nord-est de la zone sélectionnée. En comparaison avec la composition colorée, le résultat produit présente les forêts vers le sud de la zone tandis que vers le nord, il présente les forêts à la place d'une autre classe. Le produit OSO original présente plus les pelouses à la place des forêts notamment dans le sud et le centre de la zone. Une comparaison des deux produits avec la composition colorée n'a pas été faite pour les pelouses, les landes et les cultures pérennes étant donné qu'il est difficile de les différencier visuellement par photo-interprétation.

Les mêmes observations peuvent être faites pour la zone 2 où l'on remarque une occupation par les cultures pérennes et les landes dans le résultat produit (2a) vers l'est de la zone à la place des forêts dans OSO (2b). Vers le nord et le sud, les forêts ont remplacé les pelouses. En comparaison avec la composition colorée, le produit OSO d'origine (2b) présente mieux les forêts à l'est de la zone alors que le résultat obtenu (2a) montre une occupation par les landes et les cultures pérennes. Dans le nord de la zone, les deux produits présentent une répartition variable des forêts comparé à la composition colorée.

4. Discussion

Dans cette étude, nous avons expérimenté une approche méthodologique permettant d'améliorer la discrimination des surfaces des parcours du référentiel OSO 2019 en utilisant 3 différentes sources de données : les données Sentinel 2 L3A, le Référentiel parcellaire graphique de 2019 (RPG 2019) et la BD-Forêt V.2.0 actualisée en 2018.

Suite à l'entraînement de l'algorithme RF appliqué à 6 différents jeux de données d'entraînement de dimensions variables, issus du RPG 2019 et de la BD-Forêt, nous avons produit différents modèles de classification. Nous avons évalué les performances de ces modèles de classification, dans le but de définir l'approche qui permet d'obtenir les meilleurs F-scores des classes de parcours et particulièrement les landes, qui rendent compte de la présence des ligneux bas. Les résultats obtenus ont montré qu'il est difficile de mieux discriminer les landes des forêts et des pelouses avec lesquelles elles sont plus en confusion.

Le meilleur F-score des landes (0,52) a été produit à partir du jeu de données avec équilibre des effectifs d'objets pour chaque classe (50 000 par classe) intégrant l'ensemble des variables prédictives.

En effet, les landes sont très variables en termes de répartition et de densité des herbacées et des ligneux. Malgré l'intégration des variables de reflectance, d'indices spectraux et de textures (Haralick *et al.*, 1973), l'approche abordée n'a pas permis de maximiser leur séparation avec les forêts et les pelouses. En raison de cette variabilité, il a été également difficile d'améliorer leur discrimination dans OSO 2019 (Inglada *et al.*, 2017). La difficulté de mieux classer les landes est due à leurs natures très hétérogènes.

Une autre source d'erreur pourrait être les données utilisées pour l'entraînement du RF, principalement le RPG. En effet, le RPG est constitué par déclaration numérique des limites d'occupation des sols par les exploitants. Il suffit juste des débordements de limites pour que la variabilité intra-classe augmente. Cela ne reste pas sans impact pour les analyses selon l'approche par objet. Les objets utilisés pour la classification doivent être sémantiquement cohérents pour caractériser les classes qu'ils représentent (Blaschke et Strobl, 2001 ; Yu *et al.*, 2006). En vue d'améliorer cette cohérence, une application des géotraitements tels que le rétrécissement des polygones par zones tampon pourrait être expérimentée. À part les faibles précisions des landes, les modèles entraînés ont produit de meilleurs résultats pour les pelouses, les cultures pérennes, ainsi que les forêts. Leurs F-scores sont compris entre 0,83 et 0,91.

4.1. Modèles de classification

La stabilité des précisions globales respectives des 3 jeux de données utilisés montre les capacités du RF à gérer les variations dans les données d'entraînement. Le RF est moins sensible à la qualité des données d'entraînement que la plupart des algorithmes de *machine learning* (Belgiu et Dragut, 2016). Le grand nombre d'arbres de décision, la sélection aléatoire des variables et le vote majoritaire réalisé pour définir les classes (Breiman, 2001), ont permis d'optimiser l'apprentissage des modèles de classification.

Les résultats ont également montré l'influence de la taille du jeu d'entraînement sur les précisions globales et individuelles pour chaque classe. Nous avons ainsi remarqué que l'augmentation des effectifs d'objets des pelouses et des forêts favorise la diminution des précisions des landes. En effet, malgré les performances du RF à minimiser les imprécisions dues aux données non équilibrées (Khoshgoftaar *et al.*, 2007 ; More et Rana, 2017), les résultats ont montré que le niveau de confusion des classes spectralement proches peut être difficile à améliorer. Bien qu'il soit recommandé dans Millard, Richardson (2015) de réaliser un échantillonnage représentatif proportionnel des classes pour minimiser la proportion d'erreur dans la classification, aucune combinaison de nos données d'entraînement n'a pu permettre d'améliorer significativement la précision des landes.

Le modèle qui présente la grande valeur du F-score des landes (0,52) est celui qui utilise l'ensemble des variables avec des effectifs équilibrés d'objets de chaque classe.

Ce F-score reste inférieur à celui produit pour l'occupation des sols OSO 2019 qui est de 0,57. Cependant, les F-scores des pelouses et des cultures pérennes que ce modèle permet d'obtenir sont bonnes à l'échelle de nos régions d'étude. La plupart des données d'occupation des sols produites à l'échelle régionale, ne présentent pas des informations précises en ce qui concerne les surfaces pastorales. L'occupation des sols Corine Land Cover (CLC) par exemple (CGDD et SOeS, 2009), bien que présentant un grand nombre de classes, a une granularité spatiale grossière pour mieux caractériser les surfaces des parcours. Seules les surfaces supérieures à 25 ha sont représentées dans cette donnée. Or, plus de 93 % des parcelles des pelouses et plus 79 % des parcelles des landes ouvertes et fermées dans le RPG 2019 possèdent moins de 5 ha de superficie. Un grand nombre d'objets de ces classes serait contenu dans les objets individuels de CLC. En plus, la mise à jour de CLC ne se fait que chaque 6 ans, ce qui ne permet pas de suivre à court terme les évolutions des végétations pastorales, en particulier celles produites par des actions de l'homme (déroussillage, éclaircies dans les bois, etc.). L'occupation des sols OSO reste la plus homogène et précise à l'échelle nationale. Depuis 2016, l'occupation des sol OSO est mise à jour chaque année. Avec son unité minimale de cartographie de 0,1 ha, elle présente bien des avantages que CLC pour des applications au niveau des exploitations pastorales. Quelques classes sont moins discriminées dans cette source de données, en particulier les pelouses, les landes, et les cultures pérennes pour lesquelles nous avons expérimenté une approche dans le but d'améliorer leurs précisions. Cependant, l'approche utilisée dans cette étude ne permet pas d'améliorer la discrimination des landes largement utilisées dans les parcours pastoraux.

À petite échelle, Shaqura et Lasseur (2019) ont produit de meilleures précisions comparé à celles obtenues dans notre étude pour caractériser les classes pastorales. L'approche qu'ils ont développée prend en compte l'utilisation des données SPOT 6 (1,5 m de résolution spatiale en mode panchromatique et 6 m en mode multispectrale) couvrant la communauté de commune du Haut Verdon-Annot Saint-André (611 km²). Ils ont pu produire une meilleure précision globale (près de 87 % contre 83 % dans notre étude), et sont parvenus à mettre en évidence la densité des ligneux en distinguant les landes ouvertes des landes fermées. L'application des approches similaires avec ces types de données satellites à grande échelle, peuvent permettre d'obtenir des résultats similaires. Pour notre travail, il n'a pas été possible de mobiliser ce type de données pour couvrir l'ensemble de la zone d'étude d'environ 105 000 km² pour la période choisie (mi-juin à mi-juillet). Il nécessite une longue période d'acquisition (plusieurs mois) pour réunir les tuiles d'images SPOT 6/7 permettant une meilleure couverture de l'ensemble de la zone d'étude. Néanmoins, l'utilisation des informations spectrales des données Sentinel 2 L3A nous a fourni plus de variables prédictives (98 au total) afin de maximiser les possibilités de discriminer les classes pastorales.

Les résultats cartographiques (figure 7) montrent une variabilité entre l'occupation des sols produite et le produit OSO d'origine. L'approche de classification adoptée a favorisé le reclassement d'objets OSO en de nouvelles classes suite à l'entraînement des données de référence issues du RPG et de la BD-Forêt. Or, dans la plupart des cas, les

objets OSO (plus grands) contiennent plusieurs objets de référence. Cette situation détermine la grande variabilité spectrale des objets OSO pouvant être problématique pour la détermination de la classe finale.

Des différences d'occupation des sols peuvent également être perçues entre les deux occupations des sols et la composition colorée d'origine, nécessitant ainsi de nouvelles expérimentations pour améliorer l'identification des classes pastorales. Des approches beaucoup plus localisées sur de petites échelles peuvent permettre de diminuer l'effet de la variation spatiale du comportement spectral des classes de parcours et ainsi, d'accroître la discrimination des classes pastorales. Dans cette optique, des chaînes de traitement peuvent être établies pour cartographier l'occupation des sols au niveau des départements et combiner ensuite les résultats pour couvrir les régions.

4.2. Variables prédictives

Les moyennes et les écarts types de 10 bandes spectrales, 7 indices spectraux et 32 indices de texture ayant permis de constituer la liste de 98 variables prédictives ont permis d'expérimenter les possibilités de discriminer les classes pastorales. Au cours de l'entraînement du RF, un nombre réduit des variables est sélectionné au niveau de chaque arbre. Pour notre cas, 10 variables ont été aléatoirement sélectionnées pour les jeux de données intégrant toutes les variables et 5 variables pour les jeux de données à dimensionnalité réduite (soit, la racine carrée du nombre total des variables).

Les travaux de Saeys *et al.* (2007) ; Dewi et Chen (2019a) ; Dhyhan et Saurabh (2020) font référence aux techniques de réduction de dimensionnalité des données possédant un grand nombre de variables prédictives, en vue d'améliorer les performances des modèles. Il existe plusieurs techniques de réduction de dimensionnalité des variables (Saeys *et al.*, 2007). Pour ce travail, nous avons utilisé les méthodes implémentées dans le RF afin de prendre en compte l'interaction entre les données et l'algorithme de classification.

La mesure des variables importantes s'est basée sur les paramètres fixes du nombre d'arbres et des variables à utiliser au niveau de chaque nœud. Le classement des variables qui en résulte reste lié aux paramètres définis lors de l'entraînement. Ce classement pourrait être modifié en cas de modification des deux paramètres définis (Okun et Priisalu, 2007 ; Verikas *et al.*, 2011). La stabilité de cette mesure est également liée au caractère aléatoire lors de la sélection du jeu de données d'entraînement (*bagging*) (Boulesteix et Slawski, 2009 ; Genuer, 2012). Le nombre d'itération effectué (20) a permis d'évaluer la variation des scores d'importance pour chaque variable (figure 6).

Les résultats obtenus avec et sans réduction de dimensionnalité montrent des écarts faibles des précisions globales et des F-scores pour les jeux de données correspondants. Cependant, les F-scores des landes restent relativement faibles pour tous les jeux de données à dimensionnalité réduite comparé à ceux obtenus avec les jeux de données

sans réduction de dimensionnalité. Cela est sans doute dû à la complémentarité des variables non utilisées par les jeux de données à dimensionnalité réduite.

4.3. Applications et perspectives

À la suite des résultats obtenus, nous estimons que l'approche que nous avons expérimentée pourrait être améliorée en intégrant d'autres variables permettant d'accroître les possibilités de discriminer les classes pastorales. L'intégration des données temporelles pourront apporter plus d'information sur la phénologie de la végétation des parcours et ainsi donc contribuer à leur discrimination. L'utilisation des données à très haute résolution spatiale permettra également d'apporter une profondeur d'information utile pour caractériser les parcours, comme dans Shaqura et Lasseur (2019).

La disponibilité annuelle des données OSO, est une opportunité pouvant permettre de réaliser plusieurs types d'analyses spatiales (Blaschke, 2010) en vue d'améliorer la discrimination des classes pastorales. De telles analyses conserveront les caractéristiques spatiales des données OSO et présenteront une meilleure caractérisation des surfaces de parcours, ouvrant ainsi des opportunités pour des applications pastorales. Au niveau des exploitations, les éleveurs pourront s'en servir pour identifier les différentes ressources disponibles et ainsi planifier l'utilisation de l'espace par leurs troupeaux. Ce type de données permettra aussi aux gestionnaires territoriaux d'identifier la répartition des différentes surfaces de parcours pour une gestion à l'échelle du territoire.

5. Conclusion

Dans cette étude, nous avons expérimenté les potentialités offertes par les données Sentinel 2 L3A, le RPG, et la BD-Forêt V2.0, pour améliorer la discrimination des surfaces de parcours issues du Millésime OSO 2019, en zones méditerranéennes françaises. Cette expérimentation a été réalisée en utilisant l'algorithme RF qui présente un avantage d'être performant pour gérer à la fois des données non balancées et une grande quantité de nombre de variables.

L'entraînement du RF sur 6 différents jeux de données d'entraînement dont 3 avec réduction de dimensionnalité et 3 autres sans réduction de dimensionnalité, a permis d'évaluer à travers les F-scores les différentes possibilités de discriminer les classes sélectionnées. Cette approche, nous a permis de produire une meilleure discrimination des pelouses, des cultures pérennes et des forêts. Seules les landes n'ont pas été bien discriminées car étant fortement confondues avec les autres classes, à cause de leur forte variabilité en termes de types d'espèces et de leurs densités. Dans la globalité, les différents modèles ont produit de meilleures précisions avec des écarts types faibles à la suite de la validation croisée. Même si les landes n'ont pas pu être bien discriminées, la méthodologie abordée a permis de définir les éléments clés de méthode pour obtenir une discrimination satisfaisante des classes sélectionnées. Ces éléments clés sont : la

sélection d'un jeu de données équilibré (étant donné l'infériorité des effectifs des landes), l'intégration des informations spectrales, des indices de végétation, de sol et de texture comme variables prédictives, et l'entraînement par le RF d'un modèle de classification performant.

La contribution de ce travail réside dans l'utilisation des données de référence présentant des caractéristiques variables pour la classification des images satellites Sentinel 2 L3A, elles-mêmes présentant des variabilités par rapport à la précision temporelle de l'information des pixels.

À la suite des résultats de ce travail, nous estimons qu'il serait intéressant que des expérimentations intégrant l'information temporelle en plus de celles utilisées, combinées aux informations issues des données à très haute résolution spatiale telles que celles de SPOT 6/7, soient menées à de petites échelles en vue d'améliorer la discrimination des classes pastorales. Les résultats pourront être par la suite combinés pour produire une cartographie au niveau régional.

6. Annexe

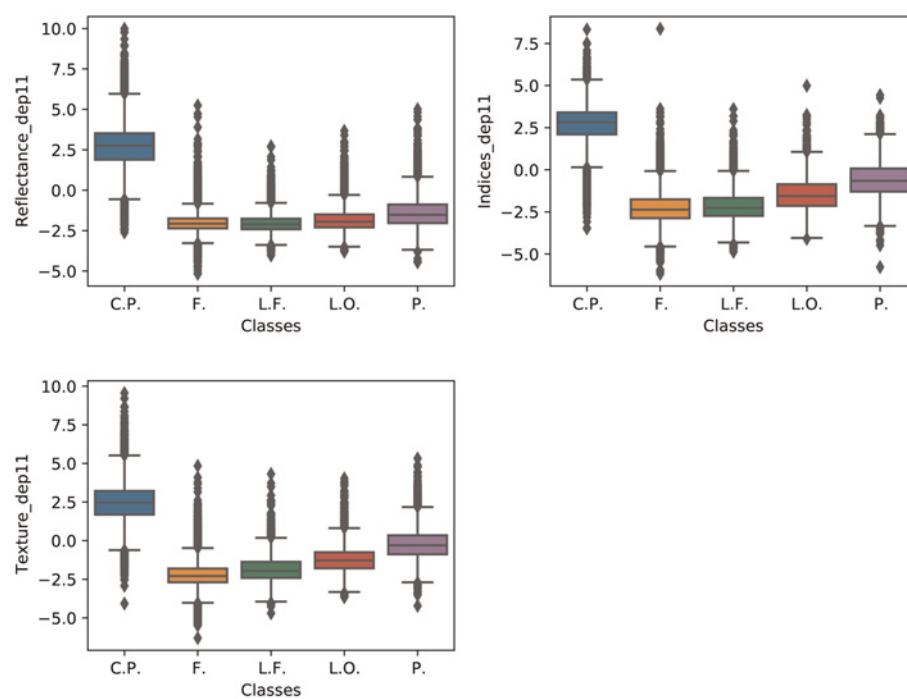


Figure 8. Boîtes à moustaches des premiers composants de l'ADL appliquée pour analyser la séparabilité des 5 classes retenues à partir des variables de reflectance, spectrales et texturales, pour le département de l'Aude (11)

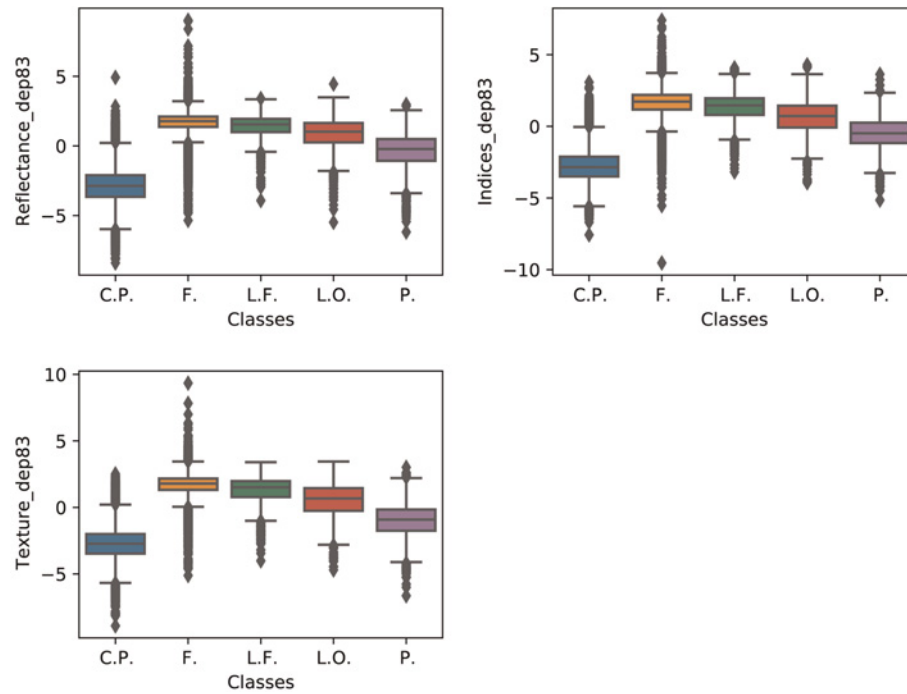


Figure 9. Boîtes à moustaches des premiers composants de l'ADL appliquée pour analyser la séparabilité des 5 classes retenues à partir des variables de reflectance, spectrales et texturales, pour le département du Var (83)

Remerciements

Ce travail a bénéficié d'une aide de l'État gérée par l'Agence nationale de la recherche au titre du programme d'Investissements d'Avenir portant la référence ANR-16-CONV-0004 (Institut Convergences Agriculture Numérique).

Bibliographie

- Akasbi Z., Oldeland J., Dengler J., Finckh M. (2012). Analysis of GPS trajectories to assess goat grazing pattern and intensity in Southern Morocco. *The Rangeland Journal*, vol. 34, n° 4, p. 415.
- Archer K.J., Kimes R.V. (2008, janvier). Empirical characterization of random forest variable importance measures. *Computational Statistics & Data Analysis*, vol. 52, p. 2249-2260.
- Belgiu M., Dragut L. (2016, avril). Random forest in remote sensing: A review of applications and future directions, vol. 114, p. 24-31.

- Benavides R., Celaya R., Ferreira L., Jáuregui B., García U., Osoro K. (2009, juin). Grazing behaviour of domestic ruminants according to flock type and subsequent vegetation changes on partially improved heathlands. *Spanish Journal of Agricultural Research*, vol. 7, p. 417.
- Blaschke T. (2010, janvier). Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 65, n° 1, p. 2-16.
- Blaschke T., Strobl J. (2001). What's Wrong with Pixels? Some Recent Developments Interfacing Remote Sensing and GIS, vol. 14, n° 6, p. 12-17.
- Bornard A., Corzic P. (2000). Les intérêts multiples des milieux pâturés d'altitude gérés par le pâturage domestique. *Le pastoralisme en France à l'aube des années 2000*. In A. Bornard, C. Brau-Nogué (Eds.), Morieres, Edition de la Cardère. (OCLC: 886107062)
- Boulesteix A.-L., Slawski M. (2009, 08). Stability and aggregation of ranked gene lists. *Briefings in Bioinformatics*, vol. 10, n° 5, p. 556-568.
- Bounejmate M., Norton B.E., Mourid M.E., Khatib A., Bathikha N., Ghassali F. *et al.* (2004). Partnership for understanding land use/cover change and reviving overgrazed rangeland in Mediterranean areas: ICARDA's experience. In A. Ferchichi (Ed.), *Réhabilitation des pâturages et des parcours en milieux méditerranéens*, p. 267-283. Zaragoza, CIHEAM.
- Breiman L. (2001). Random forests. *Machine Learning*, vol. 45, p. 5-32.
- Castro M., Ameray A., Castro J. P. (2020). A new approach to quantify grazing pressure under mediterranean pastoral systems using GIS and remote sensing, vol. 41, p. 5371-5387.
- CGDD, SOeS. (2009). CORINE Land Cover France: Guide d'utilisation.
- Colditz R. (2015, juillet). An Evaluation of Different Training Sample Allocation Schemes for Discrete and Continuous Land Cover Classification Using Decision Tree-Based Algorithms. *Remote Sensing*, vol. 7, p. 9655-9681.
- Delegido J., Verrelst J., Meza C., Rivera J., Alonso L., Moreno J. (2013, avril). A red-edge spectral index for remote sensing estimation of green LAI over agroecosystems. *European Journal of Agronomy*, vol. 46, p. 42-52.
- Derksen D., Inglada J., Michel J. (2020, février). Geometry Aware Evaluation of Handcrafted Superpixel-Based Features and Convolutional Neural Networks for Land Cover Mapping Using Satellite Imagery, vol. 12, p. 513.
- Dewi C., Chen R.-C. (2019a, octobre). Human Activity Recognition Based on Evolution of Features Selection and Random Forest. Bari, Italy, IEEE. p. 2496-2501.
- Dewi C., Chen R.-C. (2019b, décembre). Random forest and support vector machine on features selection for regression analysis. *International Journal of Innovative Computing, Information and Control*, vol. 15, n° 6, p. 2027-2037.
- Dhyan C. Y., Saurabh P. (2020, juin). Prediction of Heart Disease Using Feature Selection and Random Forest Ensemble Method. *International Journal of Pharmaceutical Research*, vol. 12.
- Drusch M., Del Bello U., Carlier S., Colin O., Fernandez V., Gascon F. *et al.* (2012). Sentinel2: Esa's optical high-resolution mission for gmes operational services. *Remote Sensing of Environment*, vol. 120, p. 25-36.

- Escadafal R. (1989, janvier). Remote sensing of arid soil surface color with Landsat thematic mapper. *Advances in Space Research*, vol. 9, p. 159-163.
- European Environment Agency. (2007). *CLC2006 Technical Guidelines*. Luxembourg, Office for Official Publications of the European Communities. Consulté sur <https://data.europa.eu/doi/10.2800/12134>
- Fawagreh K., Gaber M.M., Elyan E. (2014, décembre). Random forests: from early developments to recent advancements. *Systems Science & Control Engineering*, vol. 2, p. 602-609.
- Ferreira L.M.M., Celaya R., Benavides R., Jáuregui B.M., García U., Sofia Santos A.*et al.* (2013, août). Foraging behaviour of domestic herbivore species grazing on heathlands associated with improved pasture areas. *Livestock Science*, vol. 155, p. 373-383.
- Frenay B., Verleysen M. (2014, mai). Classification in the Presence of Label Noise: A Survey. *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, n° 5, p. 845-869.
- Genuer R. (2012). Variance reduction in purely random forests. *Journal of Nonparametric Statistics*, vol. 24, n° 3, p. 543-562.
- Haralick R. M., Shanmugam K., Dinstein I. (1973, novembre). Textural Features for Image Classification. *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-3, n° 6, p. 610-621.
- Inglada J., Vincent A., Arias M., Tardy B., Morin D., Rodes I. (2017). Operational High Resolution Land Cover Map Production at the Country Scale Using Satellite Image Time Series. *Remote Sensing*, vol. 9, n° 1, janvier, p. 95.
- Jouven M., Foulquier D., Benoit M. (2011). Graze rangelands to improve farm performance; exemple for an extensive meat sheep farming system. *Options Méditerranéennes Serie A*, vol. 100, p. 249-254.
- Jouven M., Lapeyronie P., Moulin C.-H., Bocquier F. (2010, octobre). Rangeland utilization in Mediterranean farming systems. *Animal*, vol. 4, p. 1746-1757.
- Jáuregui B. M., García U., Osoro K., Celaya R. (2009, mars). Sheep and Goat Grazing Effects on Three Atlantic Heathland Types. *Rangeland Ecology & Management*, vol. 62, p. 119-126.
- Khalilia M., Chakraborty S., Popescu M. (2011, décembre). Predicting disease risks from highly imbalanced data using random forest. *BMC Medical Informatics and Decision Making*, vol. 11, p. 51.
- Khoshgoftaar T.M., Golawala M., Hulse J.V. (2007, octobre). An Empirical Study of Learning from Imbalanced Data Using Random Forest. Patras, Greece, IEEE. p. 310-317.
- Lasanta T., Nadal-Romero E., García-Ruiz J. (2019, septembre). Clearing shrubland as a strategy to encourage extensive livestock farming in the Mediterranean mountains. *Cuadernos de Investigación Geográfica*, vol. 45, p. 487-513.
- Lin W., Wu Z., Lin L., Wen A., Li J. (2017). An Ensemble Random Forest Algorithm for Insurance Big Data Analysis, vol. 5, p. 16568-16575.
- Millard K., Richardson M. (2015, juillet). On the Importance of Training Data Sample Selection in Random Forest Image Classification: A Case Study in Peatland Ecosystem Mapping. *Remote Sensing*, vol. 7, p. 8489-8515.

- More A. S., Rana D.P. (2017, octobre). Review of random forest classification techniques to resolve data imbalance. Aurangabad, IEEE. p. 72-78.
- Okun O., Priisalu H. (2007). Random forest for gene expression based cancer classification: Overlooked issues. *Pattern Recognition and Image Analysis*, J. Martí, J. M. Benedí, A. M. Mendonça, J. Serrat (Eds.), Berlin, Heidelberg, Springer Berlin Heidelberg. p. 483-490.
- Pelletier C., Valero S., Inglada J., Champion N., Dedieu G. (2016, décembre). Assessing the robustness of Random Forests to map land cover with high resolution satellite image time series over large areas. *Remote Sensing of Environment*, vol. 187, p. 156-168.
- Perevolotsky A., Seligman N. G. (1998, décembre). Role of Grazing in Mediterranean Rangeland Ecosystems. *BioScience*, vol. 48, p. 1007-1017.
- Pouget M., Madeira E., Floch L., Kamal S. (1990). Caractéristiques spectrales des surfaces sableuses de la région côtière Nord-Ouest de l’Égypte: application aux données satellitaires SPOT. In *2^e journée de télédétection : caractérisation et suivi des milieux terrestres en régions arides et tropicales*, p. 4-6. Paris, France, ORSTOM.
- Quezel P. (1999). Les grandes structures de végétation en région méditerranéenne : facteurs déterminants dans leur mise en place post-glaciaire. *GEOBIOS*, vol. 32, n° 1, p. 19-32.
- Rancourt M. de, Fois N., Lavín M., Tchakérian E., Vallerand F. (2006, avril). Mediterranean sheep and goats production: An uncertain future. *Small Ruminant Research*, vol. 62, p. 167-179.
- Riedel J., Bernués A., Casasús I. (2013, mars). Livestock Grazing Impacts on Herbage and Shrub Dynamics in a Mediterranean Natural Park. *Rangeland Ecology & Management*, vol. 66, p. 224-233.
- Rouse J.W., Haas R.H., Schell J.A., Deering D.W. (1974). Monitoring Vegetation Systems in the Great Plains with ERTS. *NASA Special Publication*, vol. 351, p. 309.
- Saeyns Y., Inza I., Larranaga P. (2007, octobre). A review of feature selection techniques in bioinformatics. *Bioinformatics*, vol. 23, p. 2507-2517.
- Senseman G.M., Tweddale S.A., Anderson A.B., Bagley C.F. (1996). Correlation of land condition trend analysis (LCTA) rangeland cover measures to satellite-imagery-derived vegetation indices. *US Army Corps of Engineering Research Laboratories (USACERL)*, vol. 97, n° 07, p. 29.
- Shaqura I., Lasseur J. (2019, octobre). Characterizing extensively grazed areas between forests and pastureland to better assess the resilience of Mediterranean livestock activities. Lessons learned from a study on the scale of the Provence Alpes Côte d’Azur region (France). Joint Meeting FAO-CIHEAM Network on Sheep and Goats and Mediterranean Pastures “Efficiency and resilience of forage resources and small ruminant production to cope with global challenges in Mediterranean areas, 23-25 October 2019”.
- UICN. (2014). Panorama des services écologiques fournis par les milieux naturels en France. Volume 2.4: les écosystèmes montagnards. Rapport technique. France, Union Internationale pour la Conservation de la Nature.
- Van Rijsbergen C. (1979). *Information Retrieval*. (2^e édition éd.). Londres, Butterworth-Heinemann.

- Verikas A., Gelzinis A., Bacauskiene M. (2011). Mining data with random forests: A survey and results of new tests. *Pattern Recognition*, vol. 44, n° 2, p. 330-349. Consulté sur [https:// www.sciencedirect.com/science/article/pii/S0031320310003973](https://www.sciencedirect.com/science/article/pii/S0031320310003973)
- Whitcraft A. K., Becker-Reshef I., Killough B.D., Justice C.O. (2015). Meeting earth observation requirements for global agricultural monitoring: An evaluation of the revisit capabilities of current and planned moderate resolution optical earth observing missions. *Remote Sensing*, vol. 7, n° 2, p. 1482-1503.
- Yu Q., Gong P., Clinton N., Biging G., Kelly M., Schirokauer D. (2006, juillet). Objectbased Detailed Vegetation Classification with Airborne High Spatial Resolution Remote Sensing Imagery. *Photogrammetric Engineering & Remote Sensing*, vol. 72, n° 7, p. 799-811.
- Zarovali M.P., Yiakoulaki M.D., Papanastasis V.P. (2007, septembre). Effects of shrub encroachment on herbage production and nutritive value in semi-arid Mediterranean grasslands. *Grass and Forage Science*, vol. 62, p. 355-363.

Reçu le : 24 mars 2021
Accepté le : 6 décembre 2021