

# COMPLEMENTARITÉ DES IMAGES OPTIQUES SENTINEL-2 AVEC LES IMAGES RADAR (SENTINEL-1 ET ALOS-PALSAR-2) POUR LA CARTOGRAPHIE DE LA COUVERTURE VÉGÉTALE : APPLICATION À UNE AIRE PROTÉGÉE ET SES ENVIRONS AU NORD-OUEST DU MAROC VIA TROIS ALGORITHMES D'APPRENTISSAGE AUTOMATIQUE

Siham Acharki<sup>1</sup>, Pierre-Louis Frison<sup>2</sup>, Mina Amharref<sup>1</sup>, Hanaa Khoj<sup>1</sup>, Abdes Samed Bernoussi<sup>1</sup>

<sup>1</sup> Équipe GAT, Faculté des Sciences et Techniques de Tanger, Université Abdelmalek Essaadi, BP 416, Tanger, Maroc, [sacharkis@gmail.com](mailto:sacharkis@gmail.com)

<sup>2</sup> LaSTIG/MATIS, Université Gustave Eiffel, IGN, 5 Bd Descartes, Champs sur Marne, 77455 Marne la Vallée CEDEX 2, France, [pierre-louis.frison@u-pem.fr](mailto:pierre-louis.frison@u-pem.fr)

## Résumé

Dans cet article, nous évaluons les performances de classification de trois algorithmes non paramétriques (kNN, RF et SVM) en utilisant les données multi-temporelles de trois satellites (Sentinel-1, Alos-Palsar-2 et Sentinel-2) et de leurs combinaisons. La zone d'étude choisie se caractérise par un climat méditerranéen subhumide et une topographie très accidentée qui rend la classification d'occupation du sol particulièrement difficile. En outre, elle contient une aire protégée nommée Jbel Moussa et présente une diversité biologique exceptionnelle. Afin de suivre le couvert végétal de cette dernière, nous avons acquis et prétraités les images satellitaires optiques et radar pour la période du 1er janvier au 31 décembre 2017. Ensuite, nous avons combiné les trois satellites, soit douze scénarios produits. Des cartes de classifications illustrent notre approche. Un total de trente-six classifications a été obtenu, en se basant sur sept classes : eau, bâtiment et infrastructures, sol nu, végétation peu dense, prairies, forêt peu dense et forêt dense. Les résultats ont montré que pour tous les scénarios, la précision globale la plus élevée a été produite par RF (53,03%-93,06%), suivie de kNN (49,16%-89,63%), tandis que SVM (47,86%-86,08%) a produit la précision de classification la plus faible. L'étude a également montré une similitude entre les performances de la combinaison des trois satellites et celles de Sentinel-2 seul. Les estimations de la superficie pour les différentes classes vont de 0,85 km<sup>2</sup> (0,11% de la zone d'étude) à 326,84 km<sup>2</sup> (41,31% de la zone d'étude).

**Mots-clés** : Sentinel-1, Sentinel-2, Alos-Palsar-2, RF, SVM, kNN, Aire protégée

## Abstract

*In this article, we evaluate the classification performance of three non-parametric algorithms (kNN, RF and SVM), using multi-temporal data from three satellites (Sentinel-1, Alos-Palsar-2 and Sentinel-2) and their combinations. The study area selected is characterized by a subhumid Mediterranean climate and a very rough topography, making it especially difficult to classify land cover. In addition, it contains a protected area named Jbel Moussa and presents exceptional biological diversity. We have acquired and pre-processed optical and radar satellite images from January 1 to December 31, 2017, to track vegetation cover. Then to produce twelve scenarios, we combined the three satellites. Classification maps illustrate our approach. A total of thirty-six classifications were carried out, based on seven classes: water, building and infrastructure, bare soil, sparse vegetation, grasslands, sparse forest and dense forest. The results showed that the highest overall accuracy was provided for all scenarios by RF (53.03%-93.06%), followed by kNN (49.16%-89.63%), while the lowest classification accuracy was created by SVM (47.86%-86.08%). The study also showed a similarity between the performance of the three satellites combination and that of Sentinel-2 alone. Area estimates vary from 0.85 km<sup>2</sup> (0.11% of the study area) to 326.84 km<sup>2</sup> (41.31% of the study area) for various classes.*

**Keywords**: Sentinel-1, Sentinel-2, Alos-Palsar-2, RF, SVM, kNN, Protected Area

## 1. Introduction

Au cours des dernières décennies, les aires protégées ont été reconnues comme un outil essentiel pour la conservation de la biodiversité (Crabtree et Sheldon, 2011 ; Wang et al., 2020). Selon Wang et al., (2020), une aire protégée est définie comme étant une zone géographique clairement définie, acceptée, engagée et maintenue, par des moyens légaux ou autres moyens efficaces, pour assurer la protection à long terme de la nature avec les services écosystémiques et les valeurs culturelles qui lui sont associés. La gestion et la surveillance écologique des aires protégées sont indispensables pour prévenir toute perte et mettre en œuvre et évaluer la politique de leurs préservations. Toutefois, la surveillance de la biodiversité *in-situ* reste un problème dans de nombreux pays en développement, notamment dans les zones montagneuses (Vega Ishuaylas et

al., 2018). D'où la nécessité de se servir d'autres outils, tels que la télédétection, pour surmonter ce problème.

En général, la télédétection est un outil largement adopté pour la cartographie et la surveillance de l'occupation du sol, basées sur une approche de combinaison de données radar et optiques (Clerici et al., 2017 ; Denize et al. 2019 ; Acharki et al., 2020a ; Hoang et al., 2020 ; Lopez et al., 2020). Les données de télédétection multispectrales (i.e. MODIS, SPOT1/4, Landsat, Sentinel-2) fournissent des informations basées sur l'énergie diffusée (visible et infrarouge) et rayonnée (infrarouge thermique) par la surface de la terre. Par rapport aux produits dérivés des satellites précédents, les capteurs Sentinel-2 ont un contenu spatial et spectral plus riche, ainsi que des capacités opérationnelles améliorées (Clerici et al., 2017). Bien que les données multispectrales soient sensibles aux conditions atmosphériques et ne puissent être acquises que pendant la journée, elles peuvent être utilisées pour discriminer les

différentes classes de couverture terrestre en fonction des changements spectraux de leurs caractéristiques spécifiques (Nguyen et al., 2020). Cependant, les données radar à multipolarisation ou à double polarisation (i.e. Sentinel-1, PALSAR-2) constituent un bon moyen de classifier les caractéristiques homogènes de la couverture terrestre, en raison de leur capacité à pénétrer la couverture nuageuse (Haarpaintner et Heidi Hindberg, 2019 ; Refice et al., 2020). Plusieurs recherches scientifiques ont montré le potentiel de la rétrodiffusion radar en bande C et en bande L pour surveiller la couverture terrestre, en particulier le couvert forestier, induit par leur capacité à discriminer efficacement la végétation ligneuse des prairies (Borges et al., 2020 ; Frison et al., 2019 ; Haarpaintner et Heidi Hindberg, 2019 ; Hoang et al., 2020 ; Lopez et al., 2020). Il a également été démontré que les données radar en bande L (PALSAR-2 avec une longueur d'onde de 1,27 GHz, ~23,6 cm) sont généralement mieux adaptées que celles en bande C (Sentinel-1 avec longueur d'onde de 5,3 GHz, ~5,6 cm), car leur signal pénètre plus profondément dans le couvert de la forêt et capte davantage de la rétrodiffusion des branches et du tronc. Ces données fournissent ainsi plus d'informations sur la biomasse forestière (Borges et al., 2020 ; Haarpaintner et Heidi Hindberg, 2019). De plus, des études ont rapporté que les signaux à polarisation croisée (HV ou VH) ont un effet de double rebond plus faible que les signaux à co-polarisation (HH et VV) sur les zones végétalisées. D'autres études ont signalé que les données de multi-polarisation, notamment l'ajout du rapport VH/VV, améliorent la précision de la classification de la couverture terrestre (Frison et al., 2019 ; Lopez et al., 2020). Cette diversité (en longueur d'onde et en polarisation) fournit plus d'informations sur le mécanisme de diffusion en surface et contribue ainsi à discriminer les différentes classes de diffusion. Cela permet une meilleure reconnaissance des différents types de couverture terrestre (Refice et al., 2020).

À ce jour, plusieurs études ont utilisé des méthodes de classification avancées basées sur des algorithmes d'apprentissage et ont permis de constater une amélioration de la classification en combinant les données radar et optiques. Sentinel-1 et Sentinel-2 ont été utilisés par Denize et al. (2019) pour cartographier une zone agricole française en utilisant deux algorithmes non paramétriques (forêt aléatoire (RF) et machine à vecteurs de support (SVM)). Les résultats ont révélé que l'algorithme RF offrait une bonne précision en combinant les données Sentinel-1 et Sentinel-2, soit 81% avec un coefficient de kappa de 0.77. Ils suggèrent la combinaison de Sentinel-1 et -2 avec les données ALOS-2 ou TerraSAR-X. Une étude de Clerici et al. (2017) a comparé trois algorithmes d'apprentissage

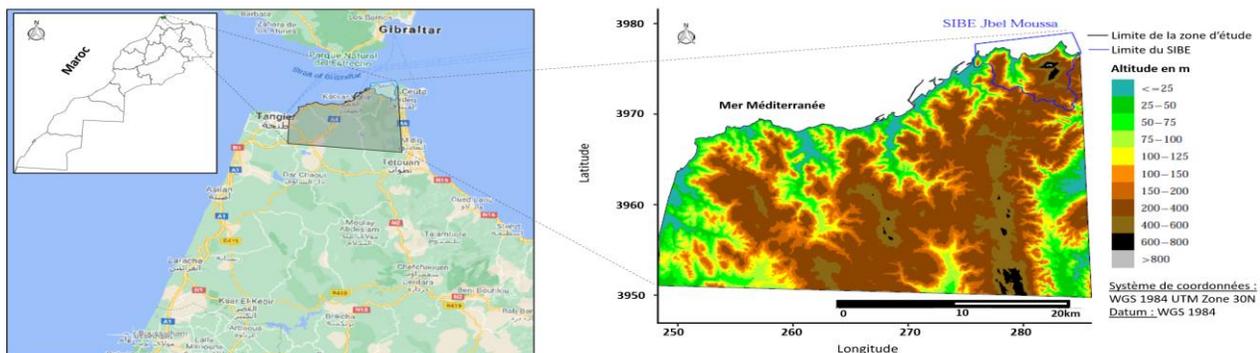
machine, à savoir le plus proche voisin (kNN), la RF et le SVM pour les données Sentinel-1 et Sentinel-2 dans le Cundinamarca-Colombie. Les résultats ont montré que le SVM était plus performant que le kNN et la RF. La combinaison a amélioré la précision globale avec une amélioration de 16,25% par rapport aux résultats obtenus avec l'utilisation de Sentinel-2 seul. Hoang et al. (2020) ont indiqué que la combinaison des images radar (PALSAR-2/ScanSAR, mosaïque PALSAR-2, Sentinel-1), optiques (Sentinel-2, Landsat-8) et des données topographiques (AW3D30), permettait de donner de bonnes performances avec une précision globale de 85,6% pour le Vietnam continental. Dans un parc du nord-ouest du Bénin, Lopez et al. (2020) ont mené une étude comparative entre Sentinel-1 et Sentinel-2. Cette étude a montré que la combinaison de ces deux satellites fonctionne aussi bien que Sentinel-2 utilisé seul, en termes de précision de la classification. D'autre part, Acharki et al. (2020a) ont constaté que l'utilisation des données Sentinel-2 seules, peut fournir des estimations raisonnables sur la cartographie des cultures pour le périmètre Loukkos-Maroc, à trois niveaux hiérarchiques (précision globale > 86%). Ces travaux ont montré que les estimations ne sont pas significativement améliorées lorsqu'elles sont intégrées à l'imagerie radar et/ou indices radiométriques.

Des recherches supplémentaires sont nécessaires pour évaluer les possibilités et les limites de l'application de la combinaison des satellites dans les régions montagneuses et dans les paysages écologiques à petite échelle. Cette étude vise à : (1) analyser la complémentarité des jeux de données optiques (Sentinel-2) et radar (Sentinel-1, Alos-Palsar-2) pour la cartographie du couvert végétal du nord-ouest du Maroc, et plus particulièrement du site à intérêt biologique (SIBE) de Jbel Moussa, qui fait partie d'une grande aire protégée transcontinentale. (2) évaluer la performance des algorithmes : Le plus proche voisin (kNN), La forêt aléatoire (RF) et La machine à vecteurs de support (SVM) pour ladite cartographie.

Cet article est organisé de la manière suivante : la section 2 présente des informations sur la zone d'étude, la section 3 décrit la démarche suivie en présentant l'ensemble de données et d'algorithmes de classification utilisés, la section 4 présente et analyse les résultats obtenus. Enfin, la section 5 présente les principales conclusions.

## 2. Zone d'étude

La zone d'étude présentée à la figure 1 est située à l'extrême nord-ouest du Maroc et couvre une superficie de 762,76 km<sup>2</sup>.



**Figure 1 :** Localisation géographique et carte d'élévation (SRTM) de la zone d'étude selon le système de projection UTM30. Nord-ouest du Maroc.

Cette zone contient une aire protégée nommée site à intérêt biologique (SIBE) de Jbel Moussa. Cette aire représente 4,6 % de la superficie totale étudiée, soit une superficie de 50 km<sup>2</sup> répartie en 35,5 km<sup>2</sup> de partie terrestre et 14,5 km<sup>2</sup> de partie marine (SPA/RAC 2019). Elle est caractérisée par une topographie très accidentée, avec des pentes abruptes, des falaises calcaires blanches et des petits replats, ce qui rend la classification d'occupation du sol particulièrement difficile. La zone étudiée se caractérise par un climat méditerranéen subhumide avec une influence océanique permanente et forte. Selon la classification de Köppen, son climat est de type tempéré chaud, avec un été sec et chaud. La température moyenne annuelle et les précipitations annuelles moyennes sont

respectivement de 17,8 °C et de 666,7 mm. Sur le plan géologique, notre zone est située à l'extrémité de la grande dorsale calcaire qui structure toute la région montagneuse du Rif. De plus, elle présente une diversité biologique exceptionnelle, liée à sa situation continentale (chaîne de Montagne du Rif) et maritime (déroit de Gibraltar). Parmi ses principaux écosystèmes, nous pouvons citer : la subéraie (Chêne-liège), les chênaies vertes, les pinèdes, les cocciferaies, les cédraies, les écosystèmes préforestiers dont les tétraclinaies. La figure 1 montre la localisation de la zone d'étude considérée et ses altitudes (SRTM) selon le système de projection UTM Zone 30N et le système de WGS 84.

### 3. Méthodologie

Le prétraitement et le traitement des images satellitaires étudiées dans cet article comportent de nombreuses étapes comme données en figure 2.

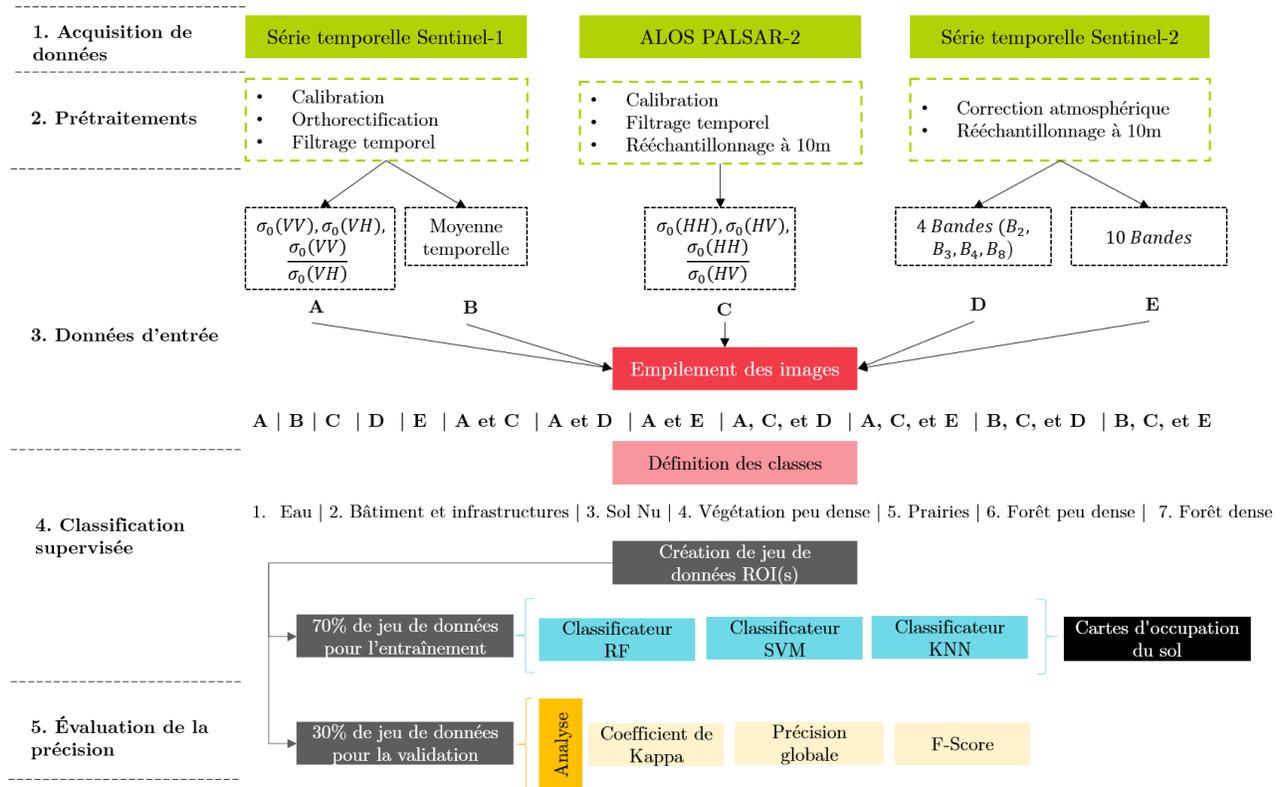


Figure 2 : Déroulement des différentes étapes suivies pour le prétraitement et la classification d'une carte de l'occupation du sol.

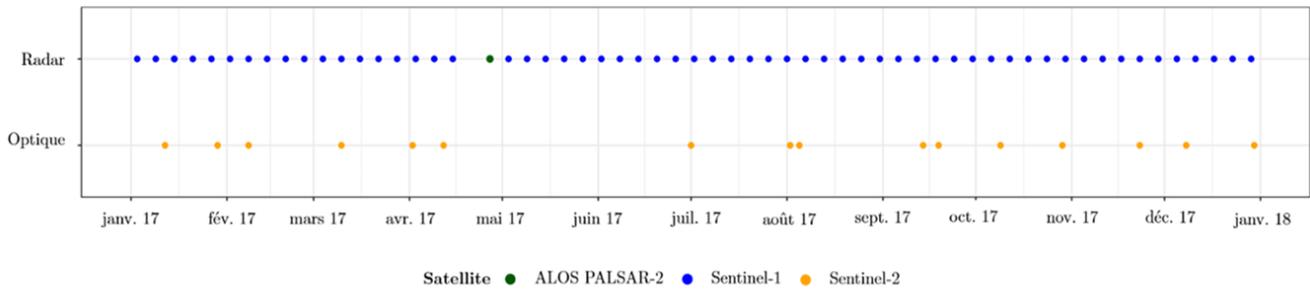
#### 3.1 Acquisition et traitement des données

Les images Sentinel-1 et Sentinel-2 ont été téléchargées à partir de la plateforme SciHub Copernicus (Copernicus, 2020) pour la période du 1er janvier 2017 au 31 décembre 2017. L'image Alos-Palsar-2, datée de 24 avril 2017 et d'une résolution de 25m, quant à elle, a été acquise via la plateforme JAXA/EORC (Jaxa, 2020) sur la zone d'étude (figure 1). L'image SRTM a été téléchargée à partir de la plateforme USGS EarthExplorer (Usgs, 2020). Elle a été utilisée pour la correction radiométrique du terrain des images Sentinel-1 et Alos-Palsar-2. Dans cet article, nous avons exploité les séries chronologiques suivantes : 60 images Sentinel-1, 16 images Sentinel-2 et une image Alos-Palsar-2. Les dates d'acquisition des images issues des trois satellites sont illustrées en figure 3.

Les images Sentinel-1 en bande C ( $\lambda \approx 5,6$  cm) comprennent des données GRD de niveau 1 en mode interférométrique large (IW). La calibration radiométrique est l'une des premières étapes importantes pour le prétraitement des images radar (Frison et Lardeux, 2018 ; Lopez et al., 2020), qui convertit les valeurs numériques en rétrodiffusion (c'est-à-dire les signaux radar renvoyés) en sigma zéro ( $\sigma_0$ ) en décibels (dB). Le calcul du coefficient de rétrodiffusion peut être effectué selon l'équation (Laur et al., 2004 ; Denize et al., 2019) [Eq. 1].

$$\sigma_{dB}^0 = 10 * \log_{10}(\sigma^0) \quad (1)$$

Après calibration, l'orthorectification (Small et Schubert, 2008) a été appliquée au jeu de données afin que celles-ci soient correctement géoréférencées. Cette étape a été réalisée



**Figure 3 :** Dates d'acquisition des images satellitaires.

en utilisant le modèle de l'ellipsoïde terrestre WGS84/UTM zone 30N.

La correction de terrain était alors appliquée pour géocoder les images. Les distorsions géométriques sont corrigées via le modèle numérique d'élévation SRTM. Cela permet de produire des images projetées au sol. Ensuite, un filtre multi-temporel (Quegan et Yu, 2001) a été appliqué aux images corrigées afin de préserver la résolution spatiale et la structure fine des images Sentinel-1. Ces traitements ont été réalisés en utilisant les scripts Python basés sur le logiciel Orfeo ToolBox (OTB, 2020). Les images résultantes en double polarisation (VH et VV) et leur rapport VH/VV ont ensuite été empilées pour produire une image moyenne temporelle, qui est utilisée pour la classification des images. Cette image temporelle permet principalement de réduire le chatolement, tout en maintenant la résolution spatiale.

Les images Sentinel-2 A et B sont de niveau 1C, ce qui signifie qu'elles ont subi une correction géométrique et radiométrique, mais pas de correction atmosphérique (Lillesand et al., 2015 ; Lopez et al., 2020 ; Nguyen et al., 2020). La correction atmosphérique, le rééchantillonnage et le découpage des bandes ont été effectués à l'aide de l'application web *Web2TISat* (Acharki et al., 2020b), implémentée sous le logiciel R (R Development Core Team, 2017). La correction atmosphérique permet de convertir les données de réflectance de la haute atmosphère de niveau 1C en un produit de réflectance de la

basse atmosphère de niveau 2A. Toutes les bandes ont été reprojétées en système WGS84/UTM zone 30N. Les bandes ayant une résolution spatiale de 20 m, ont été rééchantillonnées en 10m en utilisant l'interpolation bilinéaire, afin de faciliter l'intégration et la cohérence. Enfin, elles ont été découpées selon les limites de la zone étudiée.

L'image Alos-Palsar-2 a été acquise en bande L ( $\lambda \approx 23,6$  cm) et en double polarisation HH et HV. Elle a été construite sur un type de données de 16 bits. Tous les pixels de ces images ont des comptes numériques (DN). Ceux-ci ne représentent pourtant pas le signal radar des éléments ou des objets au sol. Par conséquent, ces DN ont été convertis en rétrodiffusion et représentés par  $\sigma^0$  en décibels (dB), en utilisant la formule suivante :

$$\sigma_{dB}^0 = 10 * \log_{10}(DN^2) - 83 \quad (2)$$

Les jeux de données produites par l'empilement des couches d'images résultantes en une image multi-bande sont décrits dans le tableau 1.

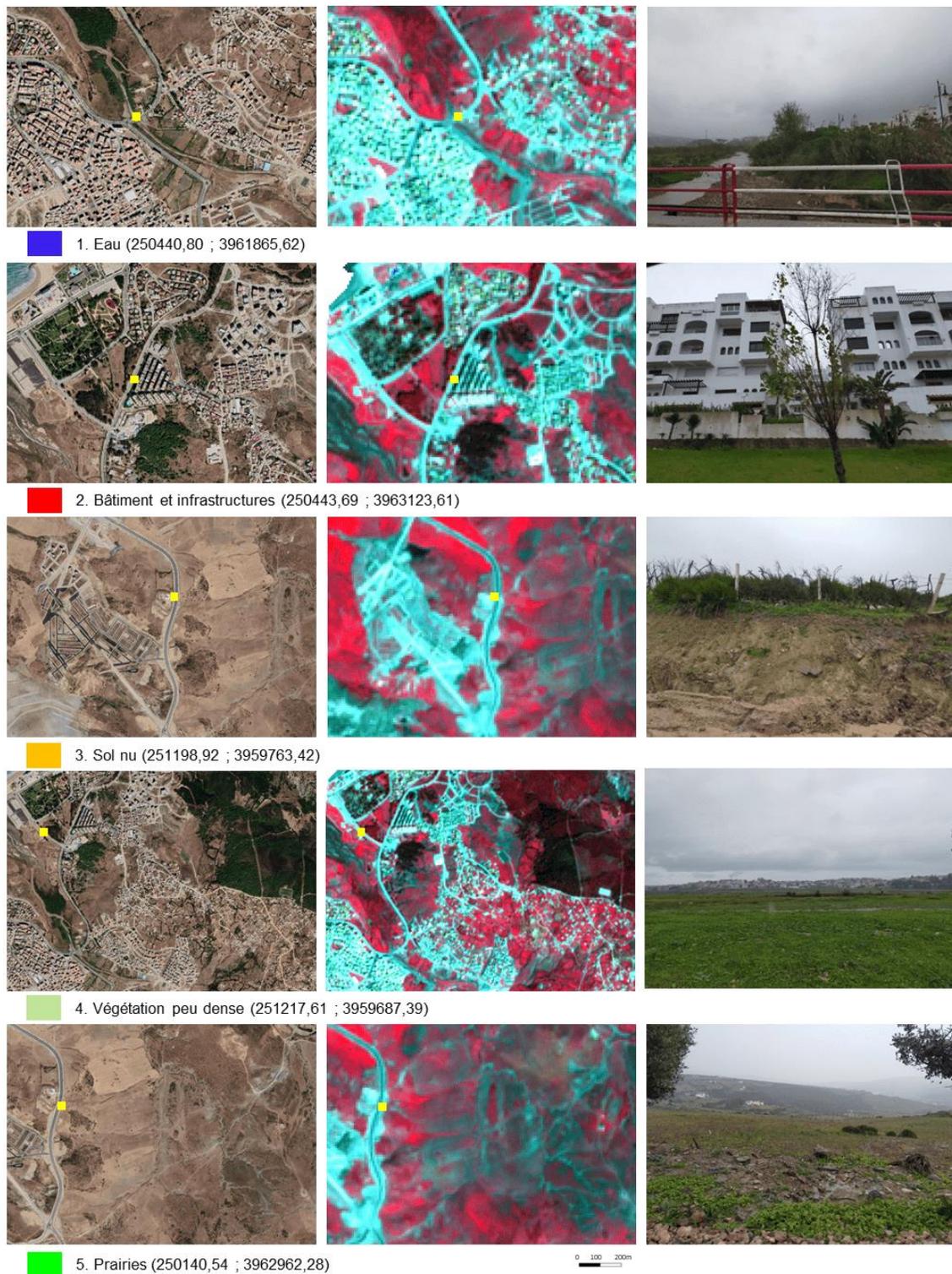
### 3.2 Classification et évaluation de la précision

#### 3.2.1. Sélection des polygones d'entraînement (ROIs)

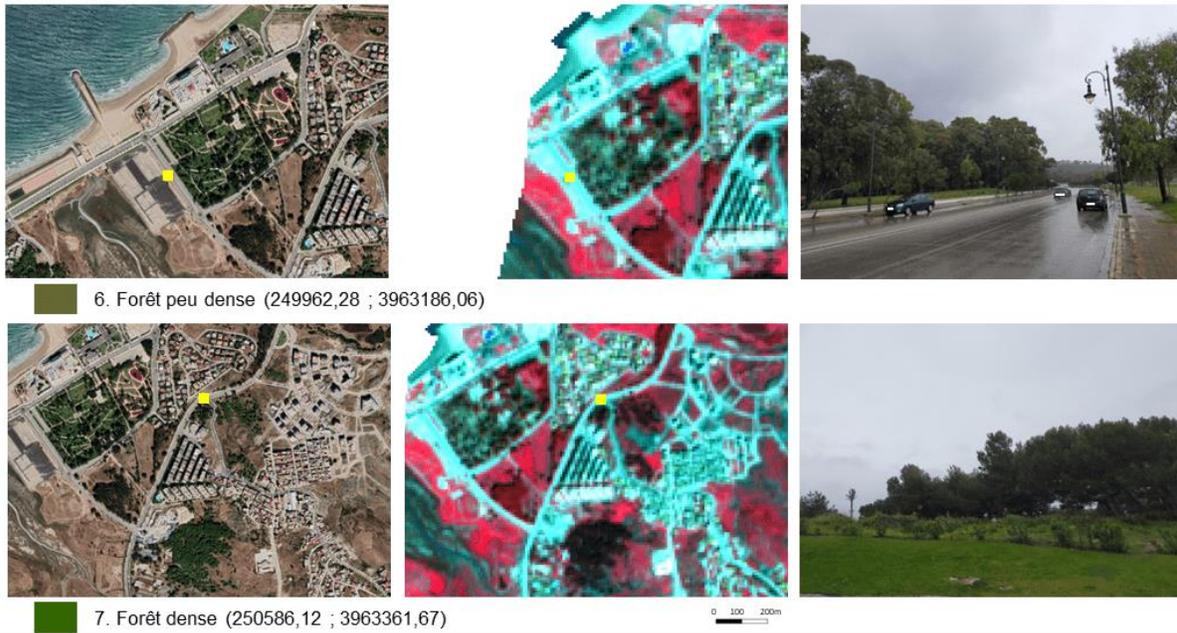
La zone d'étude est divisée en sept classes : eau, bâtiment et infrastructures, sol nu, végétation peu dense, prairies, forêt peu dense et forêt dense (figures 4 et 5).

**Tableau 1 :** Nombre de bandes des couches d'images produites.

No	Scénario	Désignation	Nombre de bandes			
			Polarisations/ Canaux	Acquisitions	Moyenne temporelle	Total
A	S1	Images Sentinel-1 : VH, VV et leur rapport VH/VV.	3	60	-	180
B	S1MT	Images Sentinel-1 et moyenne temporelle : VH, VV, leur rapport VH/VV et la moyenne temporelle.	3	60	3	183
C	P2	Images Alos-Palsar-2 : HH, HV et leur rapport HH/HV.	3	1	-	3
D	S2Q	Images Sentinel-2 : Bandes de réflectance 2, 3, 4 et 8.	4	16	-	64
E	S2B	Images Sentinel-2 : Bandes de réflectance 2 à 8A et 11 à 12	10	16	-	160
A et C	S1P2	Images Sentinel-1 et Alos-Palsar-2	3	61	-	183
A et D	S1S2Q	Images Sentinel-1 et Sentinel-2 à 4 bandes		180+64	-	244
A et E	S1S2B	Images Sentinel-1 et Sentinel-2 à 10 bandes		180+160	-	340
A, C et D	S1P2S2Q	Images Sentinel-1, Alos-Palsar-2 et Sentinel-2 à 4 bandes		180+3+64	-	247
A, C et E	S1P2S2B	Images Sentinel-1, Alos-Palsar-2 et Sentinel-2 à 10 bandes		180+3+160	-	343
B, C et D	S1MTP2S2Q	Images Sentinel-1 et moyenne temporelle, Alos-Palsar-2 et Sentinel-2 à 4 bandes		183+64	3	250
B, C et E	S1MTP2S2B	Images Sentinel-1 et moyenne temporelle, Alos-Palsar-2 et Sentinel-2 à 10 bandes		183+160	3	347



**Figure 4** : Illustration des classes d'occupation du sol : 1) eau, 2) bâtiment et infrastructures, 3) sol nu, 4) végétation peu dense et 5) prairies selon Google Earth, Sentinel-2 (02/04/2017) et correspondance sur le terrain avec les coordonnées XY en UTM Zone 30.



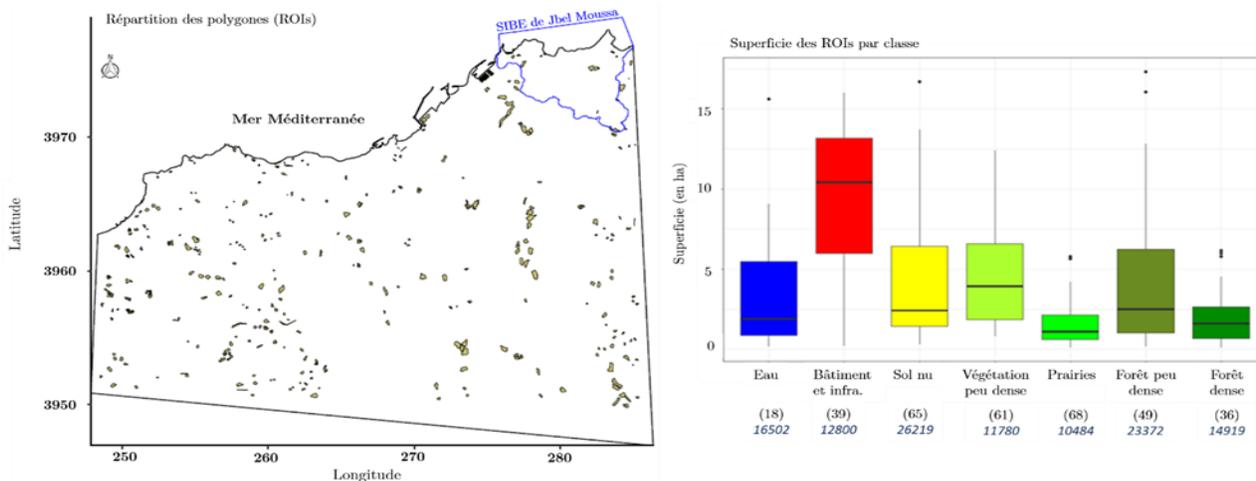
**Figure 5 :** (Suite) Illustration des classes d'occupation du sol : 6) forêt peu dense et 7) forêt dense selon Google Earth, Sentinel-2 (02/04/2017) et correspondance sur le terrain avec les coordonnées XY en UTM Zone 30.

Selon les cas, nous avons sélectionné entre 18 et 68 polygones d'entraînement pour chaque classe d'occupation des sols. Un total de 336 polygones, soit 116076 pixels, couvrait la zone en 2017. La figure 6 illustre la répartition des polygones d'intérêt utilisés (ROIs), ainsi qu'un diagramme en boîte à moustaches montrant la moyenne, les quartiles extrêmes et les déciles de chacune des classes. Nos données d'entraînement ont été réparties au hasard en deux sous-ensembles : 70 % du jeu de données ont été affectés à l'initialisation du modèle de classification, tandis que les 30% des données restantes ont été utilisées pour la validation et l'évaluation des performances du modèle. Cette répartition a été effectuée au niveau du polygone afin d'avoir un ensemble indépendant de pixels entre les étapes d'entraînement et de validation (aucun pixel appartenant au

même polygone dans les sous-ensembles d'entraînement et de validation).

### 3.2.2. Mise en œuvre des algorithmes de classification

La procédure de classification a été effectuée via le logiciel Orfeo ToolBox (OTB, 2020). Trois des algorithmes non paramétriques supervisés les plus performants ont été utilisés : Plus proche voisin (kNN : k-Nearest Neighbors), Forêt aléatoire (RF : Random Forest) et Machine à vecteurs de support (SVM : Support Vector Machine). L'un des principaux avantages de ces algorithmes est qu'ils ne nécessitent pas un modèle de distribution statistique spécifique, et qu'ils sont donc suffisamment polyvalents pour être mis en œuvre dans différents ensembles de données (Trisasonkko et al., 2017 ; Thanh Noi et Kappas, 2018).



**Figure 6 :** Répartition et superficie des polygones d'entraînement. Nord-ouest du Maroc.

Le kNN est un algorithme simple basé sur une instance connue, utilisé pour étiqueter une instance inconnue (Cover et Hart, 1967). Cet algorithme stocke toutes les données d'entraînement et classe les nouveaux points de données selon la classe de la majorité de ses  $k$  Plus proches voisins dans l'ensemble de données à disposition. Pour obtenir les voisins les plus proches de chaque donnée, kNN utilise une mesure pour calculer la distance entre des paires de données. Pour kNN, trois paramètres doivent être ajustés : la valeur maximale de  $k$  ( $k_{max}$ ), les distances des voisins les plus proches (distance) et les types de fonction du noyau (kernel).  $K$  est un paramètre de réglage clé qui influence largement les performances du classificateur kNN (Thanh Noi et Kappas, 2018) : une valeur élevée de  $k$  peut conduire à un surajustement et à une instabilité du modèle, alors qu'une valeur de  $k$  faible produira une frontière de décision très complexe. Ici, les valeurs par défaut en OTB ont été utilisées.

Le RF est un algorithme qui génère des arbres de décision en utilisant des sous-ensembles de données d'entraînement agrégées par *bootstrap*, autrement connu sous le nom de *bagging* (ensachage). La classe prédite d'une observation est calculée sur la base du vote majoritaire (classification de variable continue) ou d'une moyenne (classification par catégorie) parmi tous les arbres de décision construits dans la forêt (Breiman, 2001). Le RF requiert la sélection de trois paramètres : Ntree (nombre d'arbres à faire pousser), Mtry (nombre de variables pour diviser chaque nœud) et l'importance de la variable (le nombre de variables/bandes influence la performance du modèle) (Belgiu et Dragut, 2016). Des recherches antérieures ont étudié la sensibilité des paramètres RF et ont indiqué que parmi tous les paramètres RF, Ntree et Mtry sont les plus importants car ils peuvent avoir une influence significative sur les performances du classificateur. Belgiu et Dragut (2016) ont mis en évidence que la précision de la classification est moins sensible à Ntree qu'à Mtry. Dans cet article, nous avons utilisé les valeurs par défaut, et défini Ntree à 25 arbres et Mtry à 25 comme mentionné dans (Acharki et al., 2020a ; Acharki et al., 2020c). Selon Lopez et al., (2020), ces valeurs ont été choisies car elles ont démontré un bon compromis entre la précision de la classification et le temps de calcul.

Le SVM est une approche d'apprentissage mécanique développée sur la base de la théorie de l'apprentissage statistique et du principe de minimisation des risques structurels (Cortes et Vapnik, 1995). En général, le modèle SVM typique est séparé en deux classes, ce qui correspond au modèle d'apprentissage automatique le plus fréquemment appliqué (Ghayour et al., 2021 ; Nguyen et al., 2020), et en SVM multi-classes (combinant une chaîne de SVM à deux classes). Les performances du classificateur SVM dépendent de paramètres d'entrée définis. Selon Thanh Noi et Kappas (2018) et Nguyen et al. (2020), deux paramètres doivent être définis lors de l'application du classificateur SVM : les paramètres optimaux de coût, ou de régularisation, ou de pénalité ( $C$ ) et le paramètre de largeur du noyau ( $\gamma$ ). Le paramètre  $C$  permet de contrôler le compromis entre la maximisation de la marge et la minimisation de l'erreur d'entraînement. Le paramètre  $\gamma$  permet de lisser la forme de l'hyperplan de la division de classe. D'une part, les grandes valeurs de  $C$  entraînent des pénalités élevées, ce qui peut conduire à un surajustement. Des valeurs faibles de  $C$ , d'autre part, conduisent à un sous-ajustement, ce qui signifie que l'algorithme n'apprend pas aussi bien qu'il le pourrait

(Sonobe et al., 2017 ; Thanh Noi et Kappas, 2018). Des valeurs élevées de  $\gamma$ , en revanche, peuvent conduire à un surajustement, même si elles tentent d'ajuster les données d'apprentissage aussi précisément que possible. Tandis que les valeurs faibles de  $\gamma$  essaient de généraliser les données. Nous avons utilisé libSVM de type noyau linéaire pour mettre en œuvre l'algorithme du SVM en logiciel OTB. Le type noyau et les paramètres ont été choisis en nous appuyant sur la conclusion de Ghayour et al. (2021), laquelle établit une comparaison entre quatre types de noyaux (linéaire, polynomial, sigmoïde et à base radiale), ce qui amène les auteurs à estimer que le noyau linéaire offre la meilleure précision de classification pour Sentinel-2 et Landsat-8.

### 3.2.3. Méthodes d'évaluation de la précision

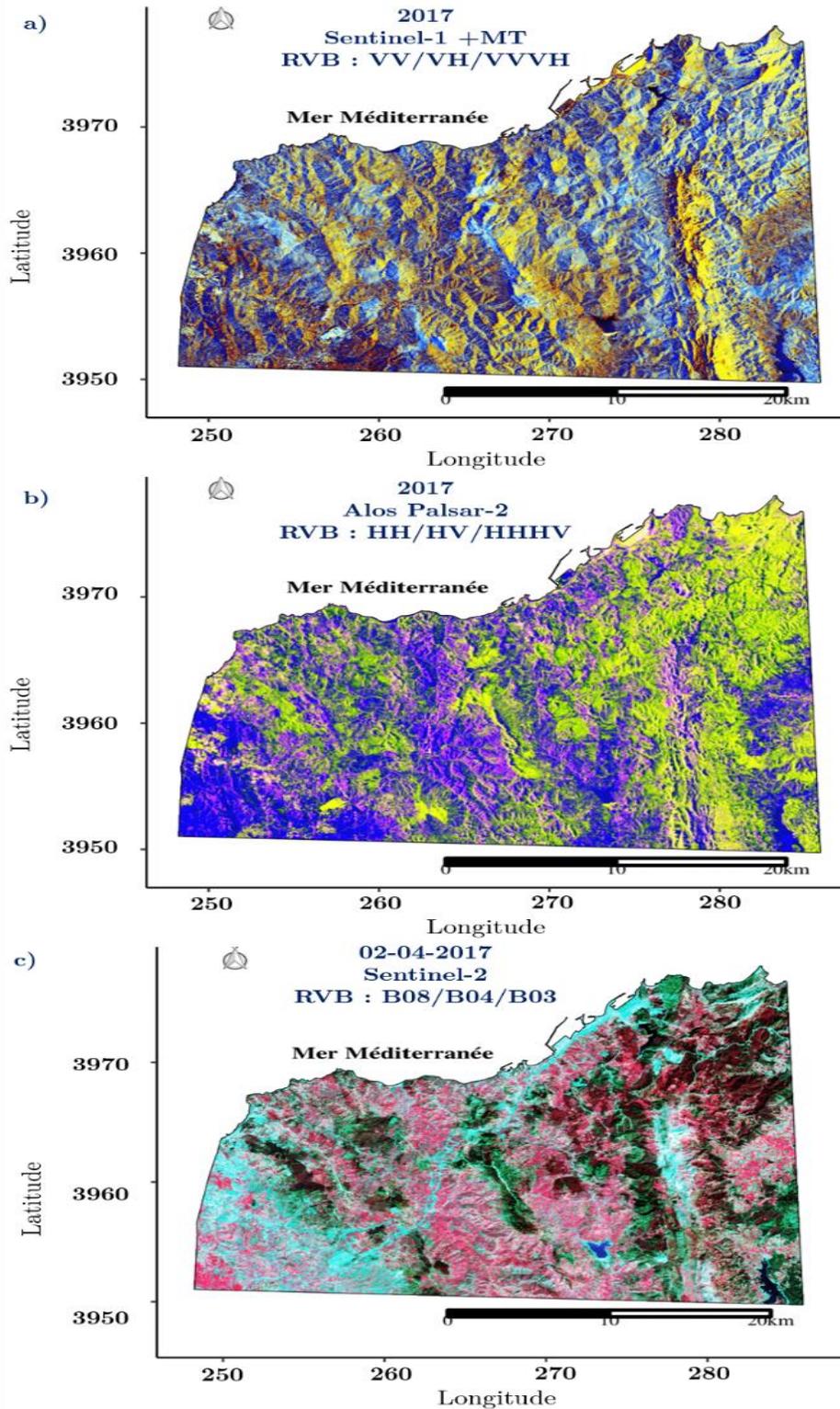
L'évaluation de la précision des résultats de la classification a été réalisée en utilisant la matrice de confusion, qui est une méthode couramment utilisée en télédétection (Lillesand et al., 2015, Nguyen et al., 2020). Pour chaque carte de classification, nous avons utilisé la précision globale et le coefficient kappa (Foody, 1992). La précision globale est calculée en divisant les pixels correctement classés par le nombre total de pixels. Le coefficient kappa, quant à lui, prend en compte l'ensemble de la matrice de confusion au lieu d'utiliser uniquement les éléments diagonaux. La précision du producteur (précision) et celle de l'utilisateur (appel) ont été utilisées pour évaluer la précision des classes individuelles. La précision est le rapport entre les pixels correctement classés d'une classe donnée, et tous les pixels de l'ensemble des données de validation pour cette classe. Le rappel est le rapport entre les pixels correctement classés dans une classe donnée, et tous les pixels classés dans cette catégorie. En outre, le F-Score (Rijsbergen, 1979) a été employé pour évaluer la pertinence du classificateur par classe. Il est calculé en utilisant la moyenne harmonique de la précision, la valeur prédictive positive et le rappel.

## 4. Résultats et Discussion

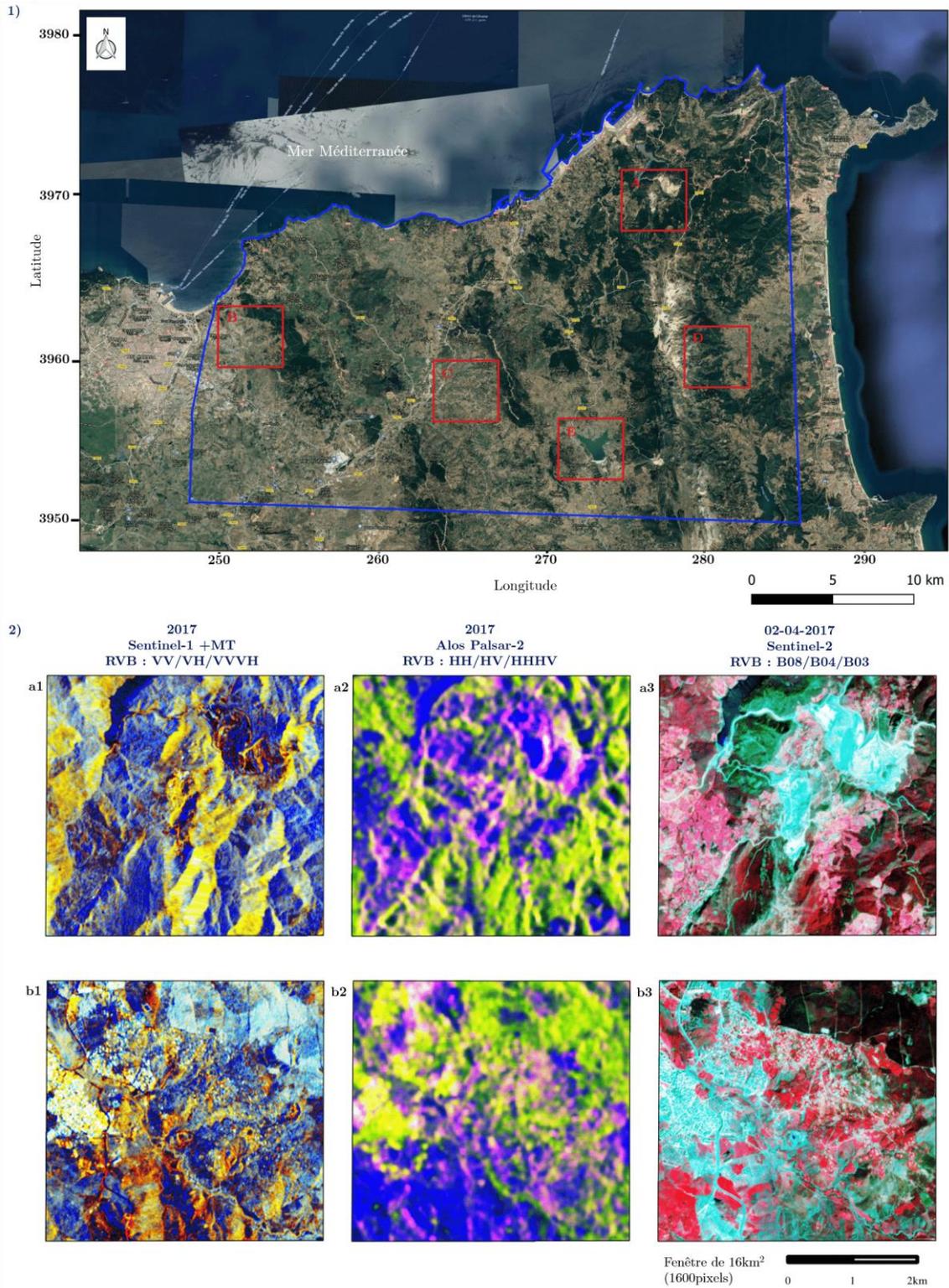
### 4.1. Analyse préliminaire

Les résultats des images composites en couleurs RVB des trois capteurs sont présentés dans la figure 7. Les interprétations visuelles de cette figure révèlent que le composite du Sentinel-2 est le meilleur et qu'il est plus informatif que les autres composites RVB.

À titre illustratif, dans la figure 8, une image Google Earth est affichée (figure 8.1) avec cinq panneaux (figure 8.2 et figure A.1) pour une meilleure interprétation visuelle. Par ailleurs, dans la figure 7.a3, nous avons une visualisation des images composites RVB sur le panneau « a », ce qui permet une analyse préliminaire mais claire de l'identification des zones forestières peu denses (couleur rose foncé) et denses (couleur rouge foncé) par rapport à d'autres types du couvert végétal. Du fait de sa grande longueur d'onde et de sa pénétration à travers la canopée, la bande L (Alos Palsar-2) ne permet la distinction ni entre les zones forestières et les prairies, ni entre les terres nues et les bâtiments et infrastructures (figure 8.a2). Aussi, la rétrodiffusion en bande C pour les sols nus ne distingue-t-elle pas des bâtiments et infrastructures (figure 8.a1). Dans l'ensemble, l'eau est la seule classe facilement distinguable dans les images composites Sentinel-1 et -2, ce qui donne un aspect bleuâtre au composite RVB.



**Figure 7 :** Résultat de la composition colorée RVB de a) imagerie Sentinel-1 avec la moyenne temporelle (B : VV, V : VH, R : VV/VH), b) imagerie Alos Palsar-2 (RVB : R : HH, V : HV, R : HH/HV) et c) imagerie Sentinel-2 datée du 02-04-2017 (R : B08, V : B04, B : B03).



**Figure 8 :** 1) Carte de localisation des panneaux rouges (Google Earth, UTM30), Nord-ouest du Maroc. 2) Composition colorée RVB des trois imageries (Sentinel-1, Alos Palsar-2 et Sentinel-2) pour les panneaux « a » et « b ».

De la même manière que pour l'exemple précédent, les sols nus et les bâtiments ainsi que les prairies et la végétation peu dense peuvent produire une rétrodiffusion similaire qui peut entraver leur distinction en bande C (figure 8.b1). En outre, aucune distinction significative des différents types de couvert végétal n'a été observée dans l'image composite Alos Palsar-2 (figure 8.b2). En revanche, l'examen visuel des résultats la figure 8.b3 montre une très bonne distinction entre les différentes classes.

#### 4.2. Évaluation de la précision globale

Les résultats de la précision globale et du coefficient de kappa pour les douze scénarios et les trois classificateurs étudiés sont présentés en tableau 2. Ces résultats montrent des performances variant entre 47,86% et 93,06%, avec des coefficients de kappa compris entre 0,41 et 0,92.

**Tableau 2** : L'indice de la précision globale et le coefficient de kappa des algorithmes : Le plus proche voisin (kNN), La forêt aléatoire (RF) et La machine à vecteurs de support (SVM) pour la classification d'occupation du sol. Nord-ouest du Maroc.

Scénario	kNN		RF		SVM	
	C <sub>ka</sub>	P <sub>0</sub>	C <sub>ka</sub>	P <sub>0</sub>	C <sub>ka</sub>	P <sub>0</sub>
S1	0,58	64,19	0,74	77,91	<u>0,39</u>	<u>47,86</u>
S1MT	0,58	63,83	0,74	78,25	0,45	51,67
P2	<u>0,41</u>	<u>49,16</u>	<u>0,45</u>	<u>53,03</u>	0,41	49,65
S2Q	0,76	79,43	0,87	89,06	0,63	67,94
S2B	0,83	85,95	0,90	91,38	0,80	82,94
S1P2	0,61	66,61	0,79	81,89	0,55	61,22
S1S2Q	0,79	82,29	0,90	91,42	0,68	72,14
S1S2B	0,87	88,75	0,91	92,36	0,76	79,48
S1P2S2Q	0,72	75,98	0,91	92,08	0,66	70,27
S1P2S2B	<b>0,88</b>	<b>89,63</b>	0,91	92,80	<b>0,83</b>	<b>86,08</b>
S1MTP2S2Q	0,75	79,44	0,91	92,04	0,73	77,69
S1MTP2S2B	0,81	83,99	<b>0,92</b>	<b>93,06</b>	0,81	84,01

V : Valeur minimale. **V** : Valeur maximale.

L'ajout de la moyenne temporelle au jeu de données Sentinel-1 permet une augmentation de la précision, soit un gain de 3,81% pour l'algorithme SVM. Tandis que pour les algorithmes kNN et RF, l'ajout ne montre pas une amélioration significative (respectivement une diminution de 0,36% et un gain de 0,34%). Il est intéressant de noter que les précisions de Sentinel-2 à dix bandes ou à quatre bandes sont supérieures à celles de Sentinel-1 et Alos-Palsar2, avec des écarts compris entre 11,15% et 38,35%. Ce résultat entre en cohérence avec les conclusions formulées par d'autres travaux (Clerici et al., 2017 ; Denize et al., 2019 ; Acharki et al., 2020a ; Hoang et al., 2020 ; Lopez et al., 2020), qui ont souligné que la classification basée sur les données Sentinel-2 a toujours été plus performante que celle basée sur les données radar. La grande sensibilité des données optiques (Sentinel-2) aux caractéristiques phénologiques des végétaux et forêts pourrait être une éventuelle explication de ces résultats (Hoang et al., 2020). En outre, les données Sentinel-2 à dix bandes fournissent de bons résultats par rapport

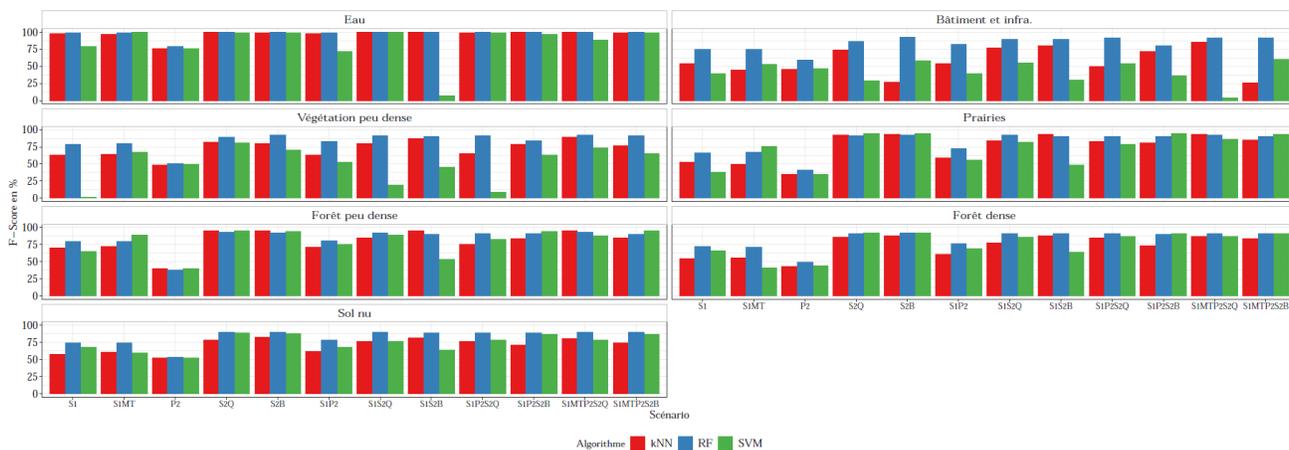
à celles à quatre bandes, correspondant à une amélioration respective de 6,52%, 2,32% et 15% pour kNN, RF et SVM. Quel que soit l'algorithme de classification, Alos-Palsar2 présente à lui seul les plus faibles précisions (<53%). Cela confirme que l'utilisation d'une seule date conduit clairement à des résultats de classification moins bons. En comparant les scénarios entre eux, nos résultats révèlent que le scénario S1P2S2B est plus performant en utilisant les algorithmes kNN et SVM, alors que pour l'algorithme RF, les scénarios S1P2S2B et S1MTP2S2B sont similaires en termes de précision globale. Globalement, pour RF, les résultats obtenus par la combinaison des jeux de données étudiés et par Sentinel-2 (dix bandes) seul ne varient pas de manière significative (écart <1,7%). Ce résultat concorde avec les travaux établis par Acharki et al. (2020a) et Lopez et al. (2020), qui ont montré une similitude entre les performances de la combinaison Sentinel-1 et Sentinel-2 et celles de Sentinel-2 seul. Cela indique que les classifications de Sentinel-2 (dix bandes) seul offrent un très bon compromis en termes de quantité de données à traiter et d'efficacité de la classification (Acharki et al., 2020a).

Dans l'ensemble, l'algorithme RF a fait preuve de la plus grande précision globale pour la combinaison des données Sentinel-1, Sentinel-2 à dix bandes ou à quatre bandes, et/ou Alos-Palsar-2 avec des précisions globales allant de 91,42% à 93,06%. Cet algorithme est suivi par kNN, avec des précisions allant de 75,98% à 89,63%, puis par SVM, pour lequel celles-ci sont comprises entre 70,27% et 86,08%. Ces résultats sont conformes à ceux obtenus par Denize et al., (2019) et Sun et al., (2019), qui ont montré que l'algorithme RF est plus précis et efficace que l'algorithme SVM, kNN ou ANN. Ces résultats sont contraires à ceux rendus par d'autres travaux (Thanh Noi et Kappas, 2018 ; Nguyen et al., 2020), pour lesquels le SVM de type noyau RBF a produit les plus grandes précisions pour la cartographie de l'occupation du sol du Vietnam, respectivement de 90% et 81%.

#### 4.2. Comparaison de la discrimination des classes

##### 4.2.1. Comparaison de l'indicateur F-Score et du pourcentage des superficies des classes

Les valeurs de F-Score pour tous les scénarios et les classificateurs sont données en figure 9. Les résultats révèlent que ces valeurs sont supérieures à 65%, à l'exception de quelques cas, mais différent selon les classes. Dans l'ensemble, la plus grande valeur moyenne de F-Score est obtenue pour Sentinel-2 (dix bandes) en utilisant l'algorithme RF, soit une valeur de 93,06%. La valeur moyenne la plus faible est obtenue pour Sentinel-1 et Sentinel-2 (dix bandes) avec l'algorithme SVM (44,52%). En ce qui concerne les scénarios combinés, les valeurs moyennes de F-Score varient de 66,73% à 89,94%, de 81,74% à 92,81%, de 44,52% à 84,46%, avec des écarts types de 6,21, 1,26 et 13,95, respectivement pour kNN, RF et SVM. Quelle que soit l'algorithme et la classe, Sentinel-2 (dix bandes) et (quatre bandes) donne de bons résultats, avec des valeurs de F-score respectivement supérieures de 70% et 73%, à l'exception de la classe bâtiment et infrastructures (26% pour kNN et 29%-57% pour SVM). Cela signifie que la majorité des classes sont bien détectées par l'optique. Pour l'algorithme RF, les différences entre les valeurs F-Score de Sentinel-2 (dix bandes) et ceux des scénarios combinés sont non significatives. En termes de F-Score, ce résultat est cohérent avec les conclusions de Acharki et al., (2020a) et Lopez et al., (2020).



**Figure 9 :** Résultats de l'indicateur F-Score selon les classes et par classificateur.

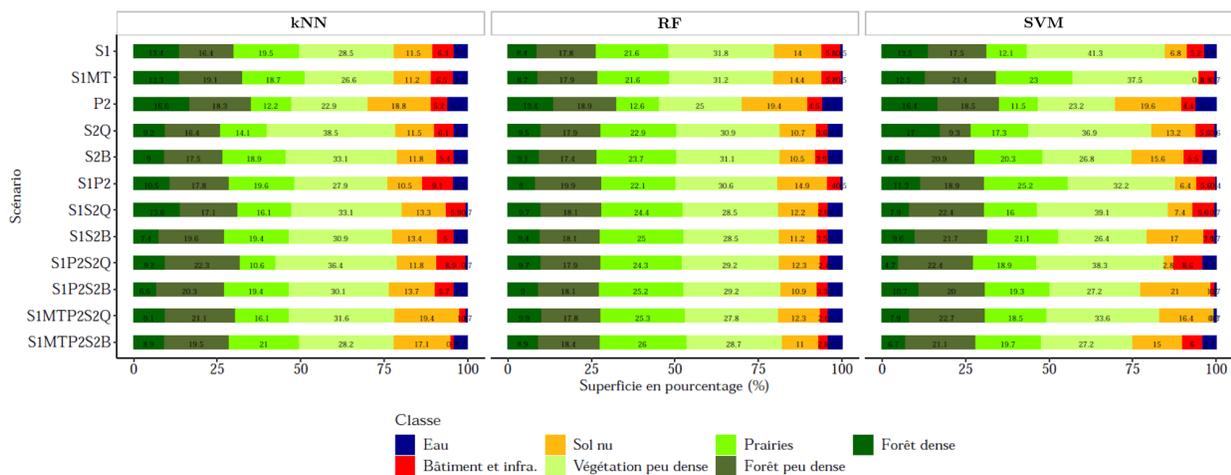
Les estimations des pourcentages de la superficie des classes d'occupation du sol sont présentées par classe pour tous les scénarios et classificateurs dans la figure 10. Ces pourcentages estimés oscillent entre 0,11% (0,85 km<sup>2</sup>) pour S1MT-SVM et 41,31% (326,84 km<sup>2</sup>) pour S1-SVM. Quel que soit le scénario, les classes ayant les plus faibles écarts-types pour RF sont respectivement forêt peu dense (5,12 km<sup>2</sup>), bâtiment et infrastructures (9,01 km<sup>2</sup>) et forêt dense (10,70 km<sup>2</sup>), tandis que, celles ayant les plus forts écarts-types sont sol nu (53,98 km<sup>2</sup>) et végétation peu dense (48,43 km<sup>2</sup>) pour SVM.

La figure A.1 illustre les distributions spatiales des classes pour le meilleur scénario et par classificateur. Les estimations de la superficie et les distributions spatiales des classes produites pourraient aider les gestionnaires pour la prise de décision et la planification de la couverture des terres forestières.

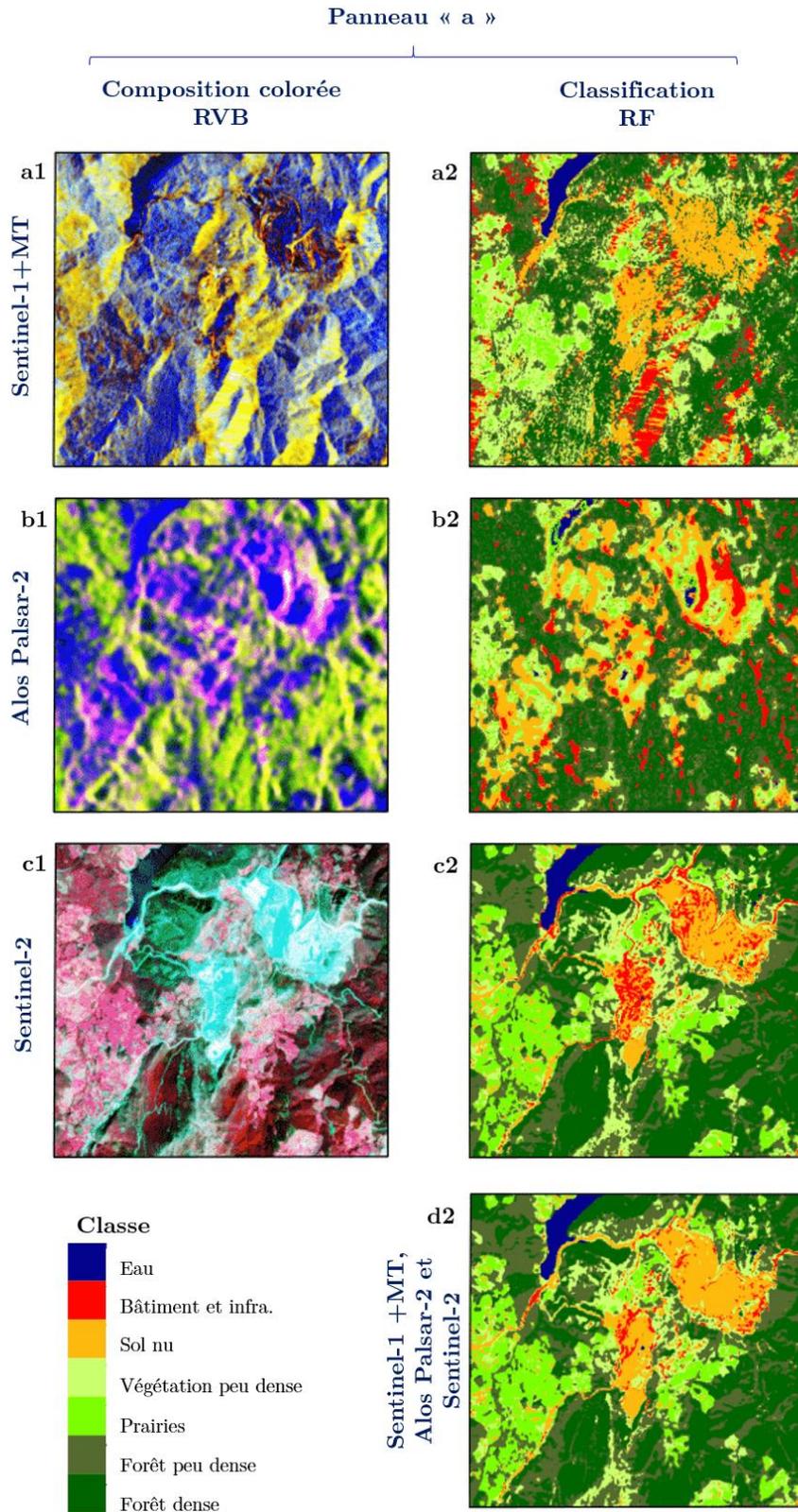
#### 4.2.1. Comparaison entre la composition colorée et la classification RF

La figure 11 illustre les images composites RVB ainsi que la classification RF des trois capteurs pour le panneau « a ». Les panneaux a1 et a2 de la figure 11 montrent que les résultats de

Sentinel-1 peuvent être assez bruyants dans les zones forestières. L'explication peut être que le radar en bande C pénètre peu dans la canopée et que sa signature représente souvent les variations de hauteur de la surface supérieure de la canopée dans les forêts mixtes non homogènes (Haarpaintner et Heidi Hindberg, 2019). Les panneaux Alos Palsar-2 (b1 et b2) ont affiché les plus mauvais résultats quant à la discrimination des différentes classes. Bien que les classes bâtiments et infrastructures et sol nu semblent mal discernées pour Sentinel-2, les résultats indiquent une bonne discrimination des différentes classes (figure 10, panneaux c1 et c2). Cette confusion pourrait être due à la ressemblance de leurs réflectances. L'utilisation des trois capteurs (figure 11, panneau d2), permet de mieux différencier les zones de couverture terrestre, en atténuant le bruit dans les zones forestières de Sentinel-1, ainsi que dans les sols nus de Alos Palsar-2. Cela permet une meilleure distinction entre prairies, forêt peu dense et forêt dense. Cette utilisation des trois capteurs permet en plus, de minimiser la confusion entre la classe bâtiments et infrastructures et la classe sol nu.



**Figure 10 :** Pourcentage de la superficie de chaque classe d'occupation du sol par scénario et par classificateur.



**Figure 11** : Comparaison de 1) la composition colorée RVB et 2) la classification supervisée par forêt aléatoire pour le panneau « a ».

## 5. Conclusion

Cette étude a évalué la complémentarité des données obtenues avec la télédétection radar ou optique pour la cartographie de l'occupation du sol dans une région montagneuse du nord-ouest du Maroc. Nous avons exploité 60 images Sentinel-1, une image Alos-Palsar-2 et 16 images Sentinel-2 pour l'année 2017. Ensuite, en utilisant trois algorithmes (kNN, RF et SVM), nous avons produit une cartographie optimale.

Nos résultats montrent que l'algorithme RF a généralement produit les meilleures performances en termes de précision globale (> 75%), de coefficient de Kappa (> 0,74) et de valeurs F-Score pour la cartographie des classes d'occupation du sol dans tous les scénarios. Ainsi, la combinaison des jeux de données Sentinel-1 et/ou Alos-Palsar-2 et Sentinel-2 peut potentiellement améliorer la précision de la classification, bien que de manière non significative. Cette étude montre également l'utilité des jeux de données Sentinel-2 pour la classification dans les zones montagneuses, et plus particulièrement dans les aires protégées à accès difficile. Les images résultantes peuvent être utilisées pour l'étude de l'effet du changement climatique, des modifications dans l'utilisation des terres, et d'autres scénarios de gestion, ainsi que pour le suivi spatio-temporel du couvert végétal et forestier. En outre, nous recommandons pour les études futures l'utilisation des séries temporelles denses, et que l'examination des espèces forestières individuelles, soit effectuée à partir de données à très haute résolution spatiale. Cela constituera en effet une grande source d'informations permettant d'améliorer les données d'inventaire forestier.

## Remerciements

*Ce travail est réalisé dans le cadre du projet de recherche PPR2/2016/79, OGI-Env, soutenu par le ministère de l'Éducation nationale, de la Formation professionnelle, de l'Enseignement supérieur et de la Recherche scientifique (MENFPESRS) et le Centre national pour la recherche scientifique et technique (CNRST). Un merci s'adresse à A. Acharki pour l'accompagnement en sorties de terrain. Un merci s'adresse également aux personnels de la Direction Régionale du Haut-Commissariat aux Eaux et Forêts et la Lutte Contre la Désertification Du Rif (HCEFLCD). Nous remercions également R. Abouelouafa (HCEFLCD) qui a participé à la réalisation de ce travail. Nous sommes reconnaissants au Pr. J-P. Rudant et aux réviseurs anonymes pour leurs précieux commentaires sur le manuscrit.*

## Références

Acharki, S., Amharref, M., Frison, P.-L., Bernoussi, A. S., 2020a. Cartographie des cultures dans le périmètre du Loukkos (Maroc) : Apport de la télédétection radar et optique. *Revue Française De Photogrammétrie Et De Télédétection*, (222), 15–29.

Acharki, S., Amharref, M., Bernoussi, A.-S., 2020b. Web2TISat: Automated satellite images downloading and processing for environmental and hydrological use. *Journal of Environmental Hydrology*, 28.

Acharki, S., 2020c. Apports de la modélisation et de la télédétection dans l'étude de l'impact des changements climatiques sur les ressources en eau : Application aux périmètres irrigués du Loukkos et du Gharb (Maroc). PhD thesis, Université Abdelmalek Essaâdi.

Belgiu, M., Dragut, L., 2016. Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing* 114, 24–31.

Borges, J., Higginbottom, T-P, Symeonakis, E., Jones, M., (2020) Sentinel-1 and Sentinel-2 Data for Savannah Land Cover Mapping:

Optimising the Combination of Sensors and Seasons. *Remote Sensing*, 12 (23), p. 3862.

Breiman, L., 2001. Machine learning. Dans: *Random Forests*. Vol. 45. Kluwer Academic Publishers, p. 5–32.

Clerici, N., Valbuena, C.A., Posada, J.M. 2017. Fusion of Sentinel-1A and Sentinel-2A data for land cover mapping: a case study in the lower Magdalena region. *Colombia, Journal of Maps*, 13:2, 718-726.

Copernicus, 2020. Copernicus Open Data Hub. Available online: URL <https://scihub.copernicus.eu>

Cortes, C., Vapnik, V., 1995. Support Vector Network. *Mach. Learn.* 20, 273–297.

Cover, T.M., Hart, P.E., 1967. Nearest neighbor pattern classification. *Ieee Transactions on Information Theory*, 13, 21.

Crabtree, R., Sheldon, J., 2011. Monitoring and modeling environmental change in protected areas: Integration of focal species populations and remote sensing. In *Remote Sensing of Protected Lands*; CRC Press: Boca Raton, FL, USA, pp. 495–524.

Denize, J., Hubert-Moy, L., Betbeder, J., Corgne, S., Baudry, J., Pottier, E., 2018. Evaluation of using sentinel-1 and -2 time-series to identify winter land use in agricultural landscapes. *Remote Sensing* 11 (1), 37.

Footy, G., 1992. On the compensation for chance agreement in image classification accuracy assessment. *Photogrammetric Engineering and Remote Sensing* 58, 1459–1460.

Frison, P.-L., Lardeux, C., Fruneau, B., Rudant, J-P., 2019. Polarimétrie Radar complète et partielle pour le suivi des surfaces terrestres. *Revue française de photogrammétrie et de télédétection*, n° 219-220 33-39.

Frison, P.-L., Lardeux, C., 2018. QGIS and Application in Agriculture and Forest. Elsevier Ltd.: Oxford, UK, Ch. *Vegetation Cartography and from Sentinel and Radar Images*, p. 181–214.

Hoang, T.T., Truong, V.T., Hayashi, M., Tadono, T., Nasahara, K. N., 2020. New JAXA High-Resolution Land Use/Land Cover Map for Vietnam Aiming for Natural Forest and Plantation Forest Monitoring. *Remote Sensing*, 12(17), 2707.

Jaxa, 2020. Centre de recherche en observation de la Terre (EORC) de l'Agence japonaise d'exploration aérospatiale (JAXA). Available online: URL <https://www.eorc.jaxa.jp/en/>

Haarpaintner J. Hindberg H., 2019. Multi-Temporal and Multi-Frequency SAR Analysis for Forest Land Cover Mapping of the Mai-Ndombe District (Democratic Republic of Congo). *Remote Sensing*, 11, 2999.

Ghayour, L., Neshat, A., Paryani, S., Shahabi, H., Shirzadi, A., Chen, W., Al-Ansari, N., Geertsema, M., Pourmehdi Amiri, M., Gholamnia, M., Dou, J., Ahmad, A., 2021. Performance Evaluation of Sentinel-2 and Landsat 8 OLI Data for Land Cover/Use Classification Using a Comparison between Machine Learning Algorithms. *Remote Sens.* 13, 1349.

Laur, H., Bally, P., Meadows, P., Sanchez, J., Schaettler, B., Lopinto, E., 2004. Derivation of the backscattering coefficient  $r$  in ESA ERS SAR PRI products. Calibration/Validation Document Issue 2, Rev. 5f, ESA.

Lee, J.-S., 1981. Speckle Analysis and Smoothing of Synthetic Aperture Radar Images. *Comput. Graph. Image Process.* 17, 24–32.

Lillesand, T., Kiefer, R., Chipman, J., 2015. *Remote Sensing and Image Interpretation*, 7th ed., Wiley: Hoboken, NJ, USA, p. 736. ISBN 978-1-118-34328-9.

Lopes, M., Frison, P., Durant, S. M., Schulte to Bühne, H., Ipavec, A., Lapeyre, V., Pettorelli, N., 2020. Combining optical and radar satellite image time series to map natural vegetation: savannas as an example. *Remote Sensing in Ecology and Conservation*, 6(3), 316–326.

Nguyen, H.T.T., Doan, T.M., Tomppo, E., McRoberts, R.E., 2020. Land Use/Land Cover Mapping Using Multitemporal Sentinel-2 Imagery and Four Classification Methods—A Case Study from Dak Nong, Vietnam. *Remote Sensing*, 12, 1367.

OTB, 2020. OrfeoToolBox. Available online: URL <https://www.orfeo-toolbox.org>

Quegan, S., Yu, J.J., 2001. Filtering of multichannel SAR images. *IEEE Trans. Geosci. Remote Sens.* 39, 2373–2379.

Refice, A., Zingaro, M., D'Addabbo, A., Chini, M. 2020. Integrating C- and L-Band SAR Imagery for Detailed Flood Monitoring of Remote Vegetated Areas. *Water*, 12(10), 2745.

R Development Core Team, 2017. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. URL <https://www.R-project.org/>

Rijsbergen, C.J.V., 1979. *Information Retrieval*. Springer-Verlag New York Inc.

Small, D., Schubert, A., 2008. Guide to ASAR geocoding. ESA-ESRIN Technical Note RSL-ASAR-GC-AD, 1.

SPA/RAC - ONU Environnement/PAM & HCEFLCD, 2019. L'aire protégée de Jbel Moussa : une perle dans le Détroit de Gibraltar. Par : Ali Aghnaj, Hocein Bazairi et Atef Limam. Ed. SPA/RAC. *Projet MedMPA Network, Tunis* : 38 pages.

Sonobe, R., Yamaya, Y., Tani, H., Wang, X., Kobayashi, N., ichiro Mochizuki, K., jul 2017. Assessing the suitability of data from sentinel-1a and 2a for crop classification. *GIScience & Remote Sensing* 54 (6), 918–938.

Sun, C., Bian, Y., Zhou, T., Pan, J., 2019. Using of multisource and multi-temporal remote sensing data improves crop-type mapping in the subtropical agriculture region. *Sensors* 19 (10), 2401.

Thanh Noi, P., Kappas, M., 2018. Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery. *Sensors*, 18, 18.

Trisasongko, B.H., Panuju, D.R., Paull, D.J., Jia, X., Griffin, A.L., 2017. Comparing six pixel-wise classifiers for tropical rural land cover mapping using four forms of fully polarimetric SAR data, *International Journal of Remote Sensing*, 38:11, 3274-3293.

Usgs, 2020. United states geological survey earth explorer. Available online : <https://www.earthexplorer.usgs.gov>

Vega Isuhuaylas, L., Hirata, Y., Ventura Santos, L., Serrudo Torobeo, N., 2018. Natural Forest Mapping in the Andes (Peru): A Comparison of the Performance of Machine-Learning Algorithms. *Remote Sensing*, 10(5), 782.

Wang, Y., Lu, Z., Sheng, Y., Zhou, Y., 2020. Remote Sensing Applications in Monitoring of Protected Areas. *Remote Sensing*, 12, 1370.

## Titre en anglais

COMPLEMENTARITY OF SENTINEL-2 OPTICAL IMAGES WITH RADAR IMAGES (SENTINEL-1 AND ALOS-PALSAR-2) FOR PLANT COVER MAPPING: APPLICATION TO A PROTECTED AREA AND ITS SURROUNDINGS IN NORTHWESTERN MOROCCO VIA THREE MACHINE LEARNING ALGORITHMS

## Légendes en anglais

Figure 1: Geographic location and elevation map (SRTM) of study area in UTM Zone 30N. Northwestern Morocco.

Figure 2: Flow chart of pre-processing and classification steps of a land cover map.

Figure 3: Satellite image acquisition dates.

Figure 4: Illustration of land cover classes: 1) water, 2) building and infrastructure, 3) bare soil, 4) sparse vegetation, and 5) grasslands according to Google earth, Sentinel-2 (02/04/2017) and field correspondence with XY coordinates in UTM Zone 30.

Figure 5: (Continued) Illustration of land cover classes 6) sparse forest and 7) dense forest according to Google earth, Sentinel-2 (02/04/2017) and field correspondence with XY coordinates in UTM Zone 30.

Figure 6: Training polygons distribution and area. Northwestern Morocco.

Figure 7: RGB color composition result of a) Sentinel-1 imagery with time average (B: VV, V: VH, R: VV/VH), b) Alos Palsar-2 imagery (RGB: R: HH, V: HV, R: HH/HV) and c) Sentinel-2 imagery dated 02-04-2017 (R: B08, V: B04, B: B03).

Figure 8: 1) Red panels location map (Google Earth, UTM30), Northwestern Morocco. 2) RGB color composition of three imageries (Sentinel-1, Alos Palsar-2 and Sentinel-2) for panels "a" and "b".

Figure 9: F-Score indicator results by class and by classifier.

Figure 10: Area percentage for each land use class by scenario and classifier.

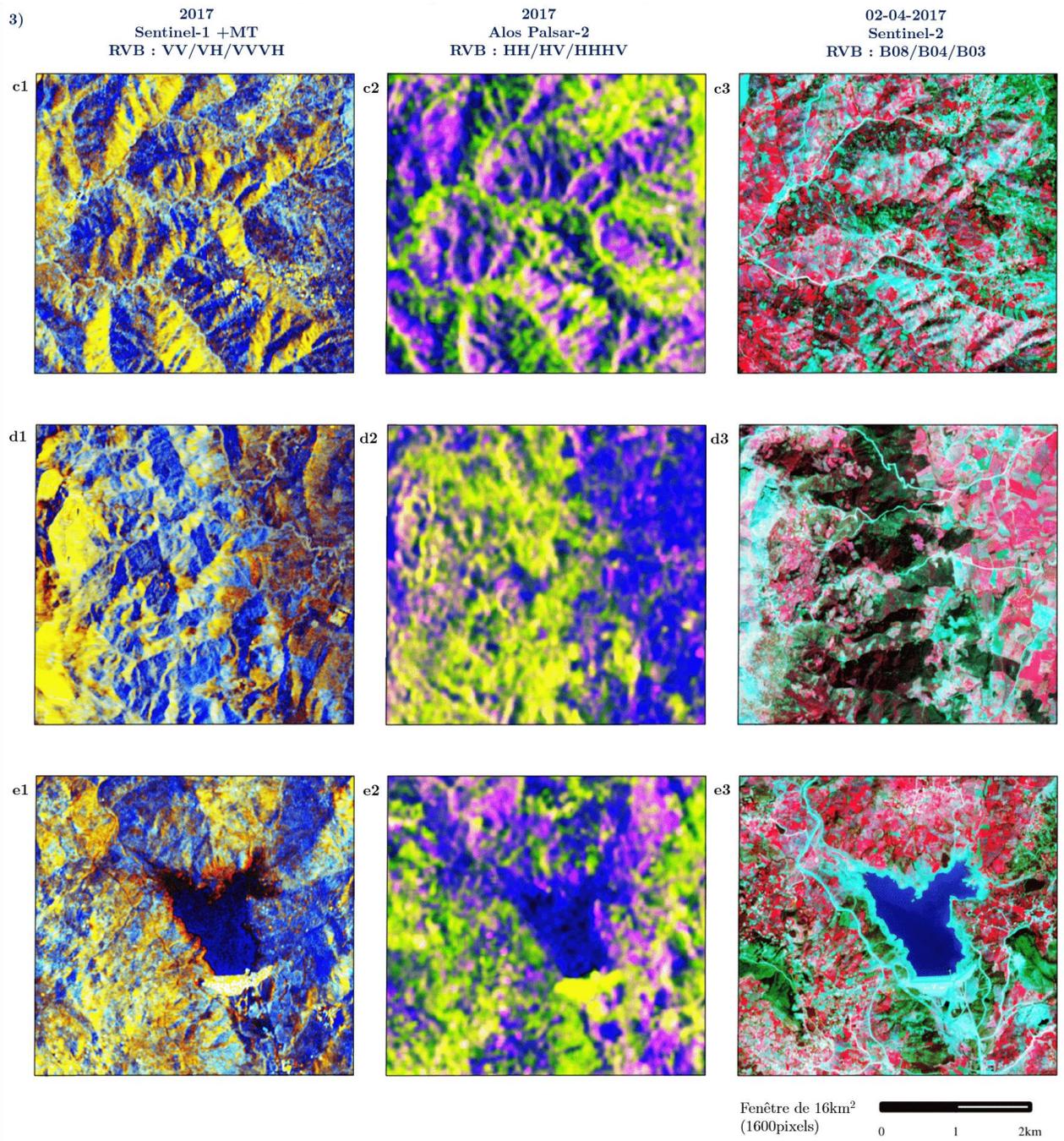
Figure 11: Comparison of 1) RGB color composition and 2) supervised random forest classification for panel "a". Figure A.1: (Continued) RGB color composition of three imageries (Sentinel-1, Alos Palsar-2 and Sentinel-2) for panels "c", "d", and "e".

Figure A.2: Land cover maps for the best scenario and by classifier. Northwestern Morocco kNN: K-nearest neighbors, RF: Random forest and SVM: Support vector machine.

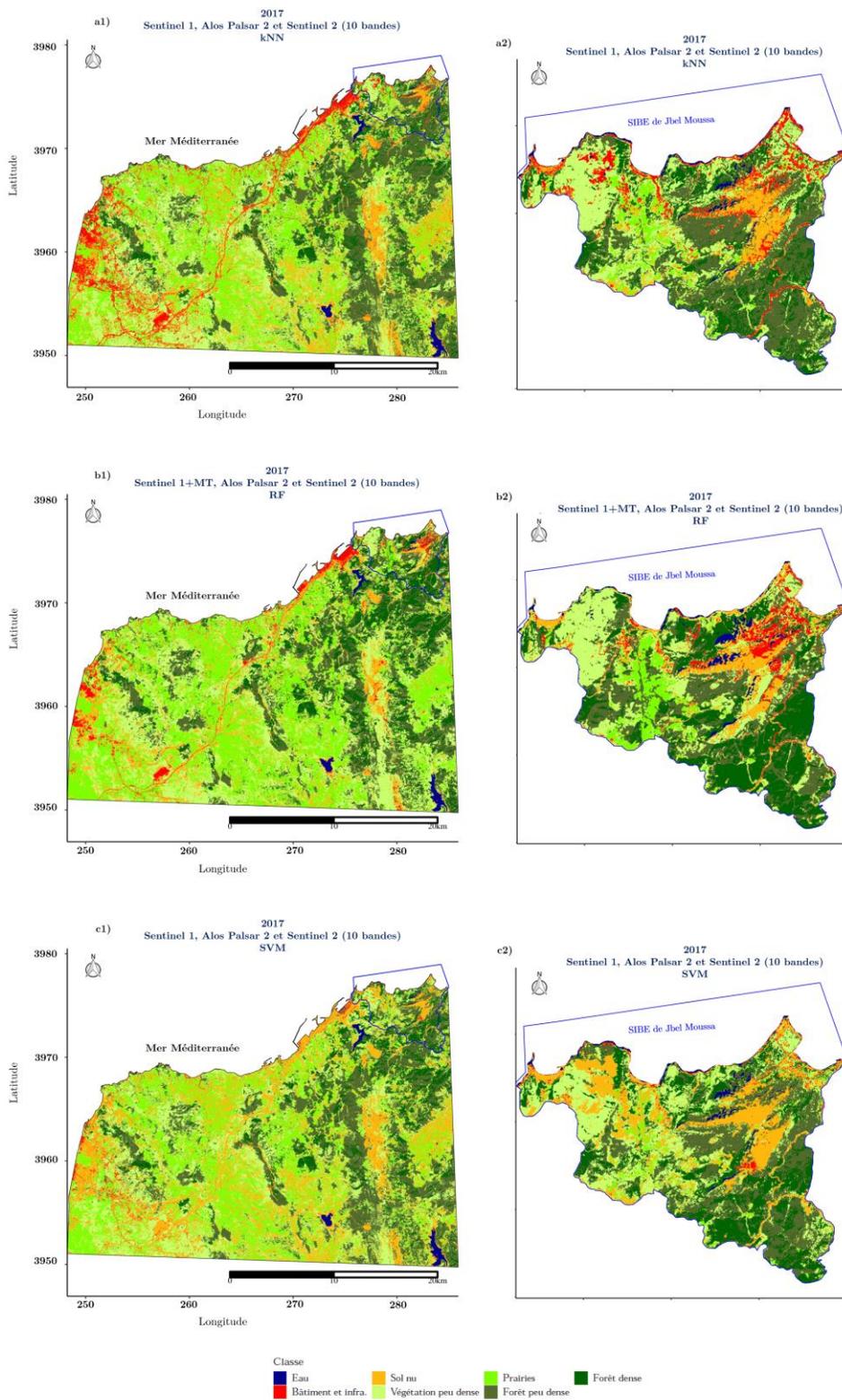
Tableau 1: Number bands of produced image layers.

Table 2: Overall accuracy index and Kappa coefficient of algorithms: k-nearest neighbor (kNN), Random forest (RF) and Support vector machine (SVM) for land cover classification. Northwestern Morocco.

Annexe A



**Figure A.1 :** (Suite) Composition colorée RVB des trois imageries (Sentinel-1, Alos Palsar-2 et Sentinel-2) pour les panneaux « c », « d », et « e ».



**Figure A.2 :** Cartes de l'occupation du sol pour le meilleur scénario et par classificateur. Nord-ouest du Maroc  
kNN : Plus proche voisin, RF : Forêt aléatoire et SVM : Machine à vecteurs de support.