Amélioration des systèmes de suivi des cultures à à l'aide de la télédétection multi-source et des techniques d'apprentissage profond

Yawogan Gbodjo

THÈSE POUR OBTENIR LE GRADE DE DOCTEUR DE L'UNIVERSITE DE MONTPELLIER

En Informatique

École doctorale : Information, Structures, Systèmes

Unité de recherche TETIS

Amélioration des systèmes de suivi des cultures à l'aide de la télédétection multi-source et des techniques d'apprentissage profond

Présentée par Gbodjo Yawogan Jean Eudes

Le 8/11/2021

Sous la direction de Ienco Dino

Devant le jury composé de

Le Saux Bertrand, Chercheur HDR, Φ-lab, ESA	Rapporteur
Mallet Clément, Chercheur HDR, LASTIG, IGN	Rapporteur
Corpetti Thomas, Directeur de recherche, LETG, Université Rennes 2	Examinateur
Forestier Germain, Professeur, IRIMAS, Université Haute-Alsace	Examinateur
Gervet Carmen, Professeur, Espace-Dev, Université de Montpellier	Examinatrice
Hubert-Moy Laurence, Professeur, LETG, Université Rennes 2	Examinatrice
Ienco Dino, Chargé de recherche HDR, TETIS, INRAE	Directeur
Leroux Louise, Chercheur, AÏDA, Cirad	Co-encadrante

UNIVERSITÉ DE MONTPELLIER

Remerciements

J'aurais aimé trouver les mots justes pour exprimer le fond ma pensée et de mes sentiments afin de témoigner ma profonde gratitude à tous ceux qui ont contribué à l'achèvement de cette merveilleuse et passionnante aventure que représente le cheminement pour l'obtention du diplôme de doctorat. L'équivocité de notre langage courant m'a obligé à chercher puis repousser à plus tard, le choix de ces mots jusqu'à ce que je réalise enfin que comme le dit l'écrivain Paolo Coehlo, « *les choses simples sont les plus extraordinaires* ».

Je voudrais ainsi tout d'abord remercier simplement les organismes qui ont financé mes travaux de thèse : IRSTEA devenu aujourd'hui INRAE et l'Institut de convergence DigitAg. À ce propos, je fais un vibrant hommage à l'équipe de direction de DigitAg qui a proposé de pertinentes activités tant intellectuelles que ludiques aux doctorants pour nous former et nous construire un réseau.

Merci infiniment à chacun des membres du jury pour l'intérêt qu'ils ont porté à ma thèse. Merci à mes rapporteurs de thèse dont l'avis critique m'a indubitablement aidé à prendre de meilleures décisions par rapport à mes travaux.

Un spécial merci à Dino Ienco, Directeur de cette thèse et Louise Leroux pour son encadrement. J'ai pris énormément plaisir à travailler avec vous ces dernières années et j'ai apprécié à leur juste valeur vos savoirs, savoir-faire et conseils qui m'ont permis de mener à terme mes recherches.

Toute ma reconnaissance au Cirad qui participe activement à la collecte de jeux de données rendues publiques auxquelles j'ai pu accéder dans la réalisation de mes travaux. Merci également à l'équipe MDL4EO, en particulier à Raffaele Gaetano et Roberto Interdonato, et à tous les collègues de TETIS pour les échanges fructueux et les moments partagés pendant ces années.

À mes parents, à ma sœur, à ma tante et aux nombreux amis qui m'ont utilement accompagné de près ou de loin tout au long de cette aventure, je vous prie sincèrement de trouver ici, la marque de ma profonde affection.

Agréable lecture.

Résumé

Les systèmes de suivi des cultures jouent un rôle essentiel dans l'évaluation de la production agricole dans le monde. De nos jours, la disponibilité de plusieurs sources d'information satellitaire à large échelle, à haute résolution spatiale et à forte répétitivité temporelle, conjointe à l'essor des techniques d'apprentissage profond, offrent de nouvelles perspectives aux systèmes de suivi des cultures pour l'évaluation de la production agricole. Dans cette thèse, nous explorons des pistes méthodologiques pour améliorer le suivi de la production agricole à partir de la télédétection multi-source et des techniques d'apprentissage profond. Nous proposons deux méthodes pour caractériser l'occupation du sol et identifier les surfaces cultivées. La première approche est basée sur des réseaux de neurones récurrents équipés de mécanismes d'attention, employant des séries temporelles multi-sources radar et optique ainsi que des connaissances spécifiques de domaine. La seconde approche repose sur des réseaux de neurones convolutifs et explore davantage la combinaison multi-source et surtout multi-échelle grâce à l'intégration d'une source optique à très haute résolution spatiale. Nous évaluons ces méthodes à des échelles territoriale et locale en ayant systématiquement un regard croisé sur des sites d'études contrastés en agriculture conventionnelle et petite agriculture familiale. Nous menons également un travail d'investigation sur l'estimation et la prévision des rendements des surfaces cultivées, à l'échelle locale de la petite agriculture familiale en employant des séries temporelles multi-sources radar et optique. Dans ce contexte en outre limité par la disponibilité de données de référence, nous évaluons le potentiel de méthodes d'apprentissage profond par rapport à des approches traditionnellement utilisées. Globalement, l'évaluation des approches proposées pour identifier les surfaces cultivées montre que les techniques d'apprentissage profond semblent mieux adaptées que les méthodes traditionnelles d'apprentissage automatique pour tirer parti de la complémentarité des données multi-sources, multi-temporelles et multi-échelles à mesure qu'il y ait une quantité suffisante de données pour leur entraînement supervisé. Le travail d'investigation réalisé pour l'estimation et la prévision des rendements n'a par contre

pas révélé de plus-value manifeste dans l'emploi de ces méthodes. Dans ce dernier cas, le contexte limité en données d'entraînement semble en être la principale explication.

Mots-clefs : Production agricole, Occupation du sol, Estimation et prévision des rendements, Données multi-sources, multi-temporelles et multi-echelles, Images radar et optique, Apprentissage supervisé, Réseaux de neurones récurrents, Réseaux de neurones convolutifs.

Abstract

Crop monitoring systems play a key role in the assessment of crop production worldwide. Nowadays, a plethora of Earth observation systems providing large scale multi-source information with high spatial and temporal resolutions, as well as the breakthrough induced by the deep learning have opened up new opportunities for crop monitoring systems in crop production assessment. This thesis investigates methodological trails in order to enhance crop production monitoring through the use of multi-source remote sensing and deep learning. We propose two methods for land cover mapping and cropland identification. The first approach is based on recurrent neural networks equipped with attention strategies which employ multi-source radar and optical time series as well as specific domain knowledge. The second approach is built on convolutional neural networks and further explores the combination of multi-source and multi-scale information especially, thanks to the integration of a very high spatial resolution optical source. We assess our proposals on territorial and local scales through a range of study sites with various landscapes, agro-climatic conditions and agricultural practices, which are located in smallholder agriculture systems and more conventional ones. We also investigate the estimation and forecasting of crop yields at the local scale of smallholder agriculture, using multi-source radar and optical time series. In this context, also characterized by a limited amount of ground reference data, we assess the potential of deep learning methods compared to commonly used modeling approaches. The evaluation of our proposals for the setting of land cover mapping and cropland identification shows that deep learning techniques seem well suited than common machine learning approaches to leverage the complementarity of multi-source, multi-temporal and multi-scale information, as there is sufficient amount of data for the training stage. On the other hand, the investigation carried out for crop yield estimation and forecasting did not show significant contributions from these methods. In this latter setting, the limited amount of ground reference data seems to be the main explanation.

Keywords: Crop production, Land cover mapping, Yield estimation and forecasting, Multi-source, Multi-temporal, Multi-scale remote sensing, SAR and optical images, Supervised learning, Deep learning, Recurrent neural networks, Convolutional neural networks.

Table des matières

Liste des figures

Liste des tableaux

Introduction

Contexte et problématique

Dans un monde en perpétuel changement fait de fortes mutations démographiques, environnementales et économiques, une augmentation substantielle de la production agricole mondiale est nécessaire pour atteindre la sécurité alimentaire globale, comme le prévoient à l'horizon 2030 les objectifs de développement durable de l'ONU (United Nations, 2015). Les systèmes de suivi des cultures jouent un rôle essentiel pour la sécurité alimentaire car ils permettent de fournir et tenir à jour des informations sur la production agricole dans le monde. Ces informations sont ainsi mises à disposition d'acteurs et de décideurs œuvrant par exemple à la prévention et à l'anticipation des pénuries alimentaires à travers des réponses politiques adaptées. À ce jour, il existe plusieurs systèmes majeurs de suivi des cultures opérationnels à l'échelle de régions, de nations ou de continents. Dans un ordre chronologique, quelques uns de ces systèmes majeurs sont : GIEWS (Global Information and Early Warning System) initié par la FAO dans les années 70, FEWS NET (Famine Early Warning Systems Network) de l'USAID établi en 1985, MARS (Monitoring Agriculture with Remote Sensing) du centre commun de recherche de la commission européenne (JRC) en 1988, CropWatch (Wu et al., 2014) introduit par l'IRSA en Chine en 1998, GEOGLAM (GEO GLobal Agricultural Monitoring) du groupe sur l'observation de la Terre (GEO) en 2013 ou encore ASAP (Anomaly Hotspots of Agricultural Production) du JRC en 2017 venu compléter MARS (Rembold et al., 2017, 2019).

La production agricole est définie comme le produit entre les surfaces cultivées et les rendements associés. Dans la plupart des systèmes existants, la cartographie des surfaces cultivées est obtenue à partir d'une synthèse de plusieurs produits globaux d'occupation du sol (Fritz et al., 2015; Waldner et al., 2016; Pérez-Hoyos et al., 2017). Cette pratique est néanmoins approximative car des contradictions importantes ont été montrées entre les divers produits utilisés (Fritz et al., 2011; Waldner et al., 2015; Pérez-Hoyos et al.,

2017). De plus, ces produits globaux réalisés à partir de données de télédétection à basse résolution spatiale n'ont pas été développés pour répondre à des problématiques agricoles et présentent de fortes incertitudes quant à la localisation des surfaces cultivées surtout en paysages agricoles complexes (Leroux et al., 2014). En définitive, elles ne permettent pas de répondre aux spécificités de toutes les régions (Waldner and Defourny, 2017). Quant à la prévision des rendements, elle est pour la plupart, non quantitative mais qualitative. Les systèmes existants donnent surtout un aperçu de l'état de développement des cultures en faisant des comparaisons avec des tendances passées afin de révéler des anomalies pouvant impacter les rendements à la récolte (Rembold et al., 2013). Cette évaluation qualitative est grandement basée sur l'analyse d'indices de végétation obtenus par télédétection. L'exemple le plus récent de système de suivi dans ce cas est ASAP (Rembold et al., 2019). Il existe donc présentement des insuffisances aux principaux systèmes de suivi des cultures dans l'estimation des surfaces cultivées et des rendements associés permettant l'évaluation de la production agricole.

La potentiel de la télédétection pour le suivi de l'agriculture a été révélé depuis plusieurs décennies (Tucker, 1979). De nombreuses études contribuant à l'évaluation de la production agricole par la cartographie de l'occupation des sols ou l'estimation des rendements se sont ainsi succédées. Les premières initiatives se fondaient surtout sur des données de télédétection optique à basse résolution spatiale comme NOAA AVHRR (1 km) et par la suite MODIS (250 m). L'avènement des satellites des programmes Landsat ou SPOT a également marqué de son empreinte ces applications en proposant des résolutions spatiales améliorées (respectivement 15/30 m et 10/20 m). La démocratisation des technologies d'observation de la Terre s'est dès lors poursuivie jusqu'à présent si bien que plusieurs sources de données satellitaires sont maintenant accessibles à large échelle sans coût pour améliorer le suivi des cultures. L'un des derniers exemples en date est celui du programme Copernicus de l'ESA avec les satellites Sentinel (Berger et al., 2012), notamment Sentinel-1 et Sentinel-2 qui fournissent des images multi-modales respectivement radar et optique de la surface terrestre à hautes résolutions spatiale (jusqu'à 10-m) et temporelle (jusqu'à 5 jours). Aujourd'hui, la télédétection multi-source représente une aubaine pour l'amélioration des systèmes de suivi des cultures au vu de la complémentarité existante et établie entre les données multi-sources (Schmitt and Zhu, 2016; Veloso et al., 2017). Toutefois, les applications existantes à l'heure actuelle sont encore majoritairement basées sur l'utilisation de données satellitaires mono-source notamment optiques.

La combinaison de données multi-source est une tâche délicate. Pour en tirer le meilleur parti, des techniques adaptées sont requises pour exploiter à la fois les spécificités des sources et leur complémentarité. De nos jours,

lorsqu'il est question d'analyse et de traitement du signal, la communauté scientifique s'oriente de plus en plus vers les techniques d'apprentissage profond (Lecun et al., 2015; Schmidhuber, 2015). Ces techniques sont aujourd'hui omniprésentes dans de nombreux domaines d'application comme la vision par ordinateur ou le traitement automatique du langage. Suite aux succès rencontrés dans ces champs, ils se sont largement exportés à d'autres domaines comme celui du suivi de l'agriculture, notamment par télédétection (Kamilaris and Prenafeta-Boldú, 2018), avec des applications en classification de l'occupation du sol (Kussul et al., 2017; Ienco et al., 2017; Rußwurm and Körner, 2018) et estimation des rendements des cultures (You et al., 2017; Khaki and Wang, 2019; Wolanin et al., 2020). La quintessence des techniques d'apprentissage profond réside dans leur capacité à découvrir automatiquement des caractéristiques optimisées pour des tâches d'apprentissage spécifiques. Les réseaux de neurones convolutifs et récurrents, deux des méthodes les plus courantes en apprentissage profond, permettent respectivement de modéliser l'auto-corrélation spatiale existante dans des images et d'analyser les dépendances au sein de données séquentielles comme des séries temporelles. En raison de leur adéquation pour le traitement d'images et de séries temporelles, les techniques d'apprentissage profond offrent aujourd'hui de nouvelles perspectives pour mieux combiner ou fusionner les données multi-sources de télédétection.

Objectifs

Au regard des insuffisances mises en évidence quant à l'évaluation de la production agricole par les systèmes de suivi des cultures, de l'opportunité offerte par la télédétection multi-source pour améliorer le suivi des cultures et des nouvelles perspectives apportées par l'apprentissage profond pour le couplage des données satellitaires multi-modales ou multi-sources, notre objectif global dans cette thèse consiste en l'exploration de pistes méthodologiques pour améliorer le suivi de la production agricole à partir de la télédétection multi-source et des techniques d'apprentissage profond.

Cet objectif de base se décline en deux objectifs spécifiques : caractériser l'occupation du sol pour identifier entre autres les surfaces cultivées ainsi qu'estimer et prévoir les rendements sur ces surfaces cultivées. Indépendamment de ces objectifs spécifiques, nous nous posons les questions de recherche suivantes :

— les techniques d'apprentissage profond seront-elles en toute circonstance meilleures que les méthodes classiques en apprentissage automatique et ceci sur des sites d'étude contrastés localisés à la fois en

agriculture conventionnelle et petite agriculture familiale ?
— quel est le comportement des méthodes d'apprentissage profond vis-à-vis de données d'apprentissage limitées ce qui est le propre des scénarios plus ou moins opérationnels ?

Pour répondre à notre objectif de base ainsi que nos questions de recherche, nous avons ainsi proposé pour la caractérisation de l'occupation du sol, en un premier temps, une méthode basée sur des réseaux de neurones récurrents employant des séries temporelles multi-sources radar et optique. Ensuite, nous avons proposé une seconde méthode basée sur des réseaux de neurones convolutifs et exploré davantage la combinaison multi-source et surtout multi-échelle par l'ajout d'une source optique à très haute résolution spatiale aux séries temporelles précédentes. Dans notre démarche, nous avons travaillé à la fois à des échelles territoriale et locale en ayant systématiquement un regard croisé vis-à-vis de l'application de nos méthodes sur des sites d'étude contrastés par leurs paysages, leurs conditions agro-climatiques et leurs pratiques culturales. Ces sites sont localisés à la fois en agriculture conventionnelle et en petite agriculture familiale.
Quant à la modélisation des rendements sur les surfaces cultivées, identifiées grâce à la caractérisation de l'occupation du sol, nous avons traité d'une application à la culture du mil dans le cas de la petite agriculture familiale en Afrique subsaharienne. Dans un contexte où les données d'apprentissage disponibles sont très limitées, nous avons réalisé un travail d'investigation sur l'estimation et la prévision des rendements du mil à une échelle locale à partir de séries temporelles multi-source radar et optique tout en évaluant le potentiel de méthodes d'apprentissage profond par rapport à des approches traditionnellement utilisées.

Plan de la thèse

Cette thèse est subdivisée en quatre chapitres après cette introduction.
Le chapitre 1 pose des notions de base pour une compréhension optimale des travaux réalisés. Nous y reportons dans une première section les principes fondamentaux de la télédétection aussi bien optique que radar et décrivons les programmes spatiaux constituant les sources d'images satellitaires utilisées tout au long des recherches menées. Dans une seconde section, nous fournissons les éléments principaux de l'apprentissage automatique en nous focalisant sur les techniques d'apprentissage profond auxquelles nous avons eu recours dans nos travaux.
Le chapitre 2 est quant à lui dédié à la présentation des données utilisées dans nos travaux. Nous présentons en premier nos sites d'études, leurs spécificités

ainsi que les données qui y sont collectées. Vient ensuite la partie consacrée à la description des prétraitements réalisés sur ces données en l'occurrence ceux effectués sur les images satellitaires.
Le contenu des chapitres suivants (3 et 4) est au cœur des travaux menés dans cette thèse. Les deux chapitres viennent apporter des pistes de réponse à notre objectif général et s'articulent autour de trois articles scientifiques :

[Gbodjo et al., 2020] Y. J. E. Gbodjo, D. Ienco, L. Leroux, R. Interdonato, R. Gaetano, and B. Ndao, "Object-Based Multi-Temporal and Multi-Source Land Cover Mapping Leveraging Hierarchical Class Relationships," in *Remote Sensing*, vol. 12, no. 17, p. 2814, Aug. 2020, doi:10.3390/rs12172814.

[Gbodjo et al., 2021a] Y. J. E. Gbodjo, O. Montet, D. Ienco, R. Gaetano and S. Dupuy, "Multi-sensor Land Cover Classification With Sparsely Annotated Data Based on Convolutional Neural Networks and Self-Distillation," in IEEE *Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 11485-11499, 2021, doi:10.1109/JSTARS.2021.3119191.

[Gbodjo et al., 2021b] Y. J. E. Gbodjo, D. Ienco and L. Leroux, "Benchmarking statistical modelling approaches with multi- source remote sensing data for crop yield monitoring in Sub-Saharan Africa," in *International Journal of Remote Sensing*, 42 :24, 9285-9308, doi:10.1080/01431161.2021.1993465

Le chapitre 3 est lié aux deux premiers articles. Il traite de la caractérisation de l'occupation du sol pour identifier les surfaces cultivées à partir des techniques d'apprentissage profond et de la télédétection multi-source. Nous reportons dans deux sections successives nos différentes contributions méthodologiques : tout d'abord celles du premier article puis du second. Le fonctionnement de chaque méthode est tout d'abord présenté dans la section dédiée. Ensuite, le protocole expérimental d'évaluation englobant les sites d'étude, les données utilisées et les approches de comparaison, est décrit. Enfin les résultats de l'évaluation sont présentés et discutés.
Le chapitre 4 fait référence au troisième et dernier article cité. Il traite de l'estimation et de la prévision des rendements avec une application à la culture du mil dans le cas de la petite agriculture familiale en Afrique subsaharienne. Nous y évaluons le potentiel des techniques d'apprentissage profond par rapport aux approches traditionnelles à travers la combinaison multi-source dans un contexte très limité en données d'apprentissage.
En dernier lieu, nous concluons ces travaux de thèse en résumant leurs principales contributions et ouvrons sur quelques perspectives du point de vue

de leur intégration aux systèmes de suivi des cultures.
Nous avons également communiqué sur ces travaux de thèse lors de colloques nationaux et actes de conférences internationales.

- Y. J. E. Gbodjo, D. Ienco, L. Leroux, R. Interdonato, R. Gaetano and B. Ndao, "RNN-based Multi- Source Land Cover Mapping : Application to a West African Agricultural Landscape," in 1^{th} *Symposium of GdR MADICS*, June 2019, Rennes, France.

- Y. J. E. Gbodjo, L. Leroux, R. Gaetano and B. Ndao, "RNN-based Multi-Source Land Cover mapping : An application to West African landscape," in *MACLEAN Workshop ECML PKDD 2019*, September 2019, Wurzburg, Germany.

- Y. J. E. Gbodjo, D. Ienco, L. Leroux, R. Interdonato, and R. Gaetano, "Fine grained classification for multi-source land cover mapping," in *Computer Vision for Agriculture Workshop ICLR 2020*, April 2020 Addis-Abeba, Ethiopia.

En marge des articles auxquels sont associés les travaux de cette thèse, nous avons aussi mené d'autres recherches toujours en lien avec la classification de l'occupation du sol et l'estimation des rendements qui ont donné lieu à d'autres publications dans des revues scientifiques à comité de lecture :

- Y. J. E. Gbodjo, D. Ienco and L. Leroux, "Toward Spatio–Spectral Analysis of Sentinel-2 Time Series Data for Land Cover Mapping," in *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 2, pp. 307-311, Feb. 2020,doi:10.1109/LGRS.2019.2917788.

- D. Ienco, Y. J. E. Gbodjo and R. Gaetano, "Generalized Knowledge Distillation for Multi-Sensor Remote Sensing Classification :An Application to Land Cover Mapping," in *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2020, vol. 5, p. 997-1003, doi:10.5194/isprs-annals-V-2-2020-997-2020.

- D. Ienco, Y. J. E. Gbodjo, R. Gaetano and R. Interdonato, "Weakly Supervised Learning for Land Cover Mapping of Satellite Image Time Series via Attention-Based CNN," in *IEEE Access*, vol. 8, pp. 179547-179560, 2020, doi:10.1109/ACCESS.2020.3024133.

- A. M. Censi, D. Ienco, Y. J. E. Gbodjo, R. pensa and R. Gaetano, "Attentive Spatial Temporal Graph CNN for land cover mapping from multitemporal remote sensing data," in *IEEE Access*, vol. 9, pp. 23070-23082, 2021, doi:10.1109/ACCESS.2021.3055554.

- L. Leroux, G.N. Falconnier, A.A. Diouf, B. Ndao, Y. J. E. Gbodjo, L. Tall, A.A. Balde, C. Clermont-Dauphin, A. Bégué, F. Affholder, O. Roupsard, "Using remote sensing to assess the effect of trees on millet yield in complex parklands of Central Senegal," in *Agricultural Systems*, vol. 184, 2020, 102918, doi:10.1016/j.agsy.2020.102918.

Chapitre 1

Notions préliminaires : Télédétection et Apprentissage automatique

Sommaire

1.1 Télédétection

La télédétection désigne un ensemble de techniques permettant la mesure à distance de propriétés physiques caractéristiques d'objets étudiés. Ces techniques englobent plusieurs étapes qui vont de l'acquisition de la mesure jusqu'au traitement et l'analyse de cette information. Quoi qu'il en soit, ce processus met en œuvre un ou plusieurs capteurs embarqués sur une plate-forme mobile. Les capteurs sont du type caméra, radar, laser ou sonar tandis que les plate-formes peuvent être des satellites, avions, navires ou même drones. Ainsi, en fonction des capteurs et plate-formes utilisés, nous distinguons plusieurs types de télédétection. Dans cette thèse, nous nous intéressons principalement à la télédétection spatiale notamment par les satellites d'Observation de la Terre (OT).

Le principe de base de la télédétection repose sur la mesure du rayonnement émis ou réfléchi par les objets dans différents domaines du spectre électromagnétique (Figure 1.1). L'exploitation des domaines du visible et de l'infrarouge du spectre électromagnétique définit les systèmes de télédétection optique tandis que celui des micro-ondes définit les systèmes de télédétection radar.

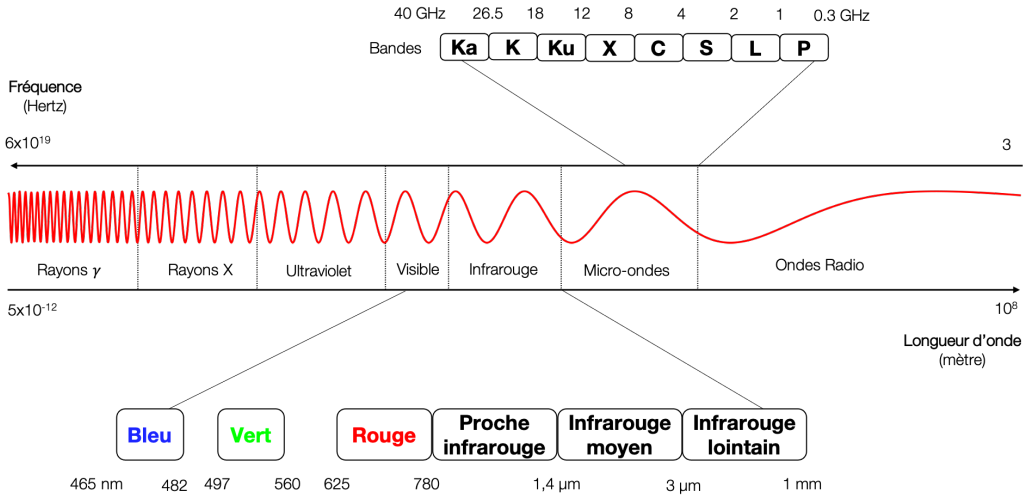


Figure 1.1 – Les différents domaines du spectre électromagnétique

1.1.1 Satellites optiques

Les satellites optiques d'OT emploient le plus souvent des radiomètres multi-spectraux comme capteurs. Ces derniers sont dits passifs car ils ex-

ploitent le rayonnement électromagnétique ayant pour source le soleil. Ces capteurs reçoivent et mesurent une partie du rayonnement solaire réfléchi ou absorbé puis émis en leur direction par l'atmosphère et les objets de la surface terrestre. Les capteurs optiques analysent divers domaines du spectre électromagnétique, couramment : le visible, le proche infrarouge, l'infrarouge thermique et le moyen infrarouge. Cette capacité à analyser la réponse spectrale [1] des objets dans plusieurs longueurs d'onde du spectre électromagnétique définit la richesse ou résolution spectrale des satellites optiques. Cette richesse spectrale met en exergue des propriétés caractéristiques des surfaces sur lesquelles se fondent notamment les indices spectraux de télédétection. Un exemple notoire est celui de la végétation qui absorbe très fortement les longueurs d'ondes du rouge et réfléchit tout autant dans la même proportion les longueurs d'ondes du proche infrarouge. Nous détaillons plus bas quelques indices spectraux de télédétection notamment des indices de végétation.

Un autre aspect particulièrement important des satellites d'OT en général réside dans les résolutions spatiale et temporelle qui les caractérisent. Le premier définit l'emprise au sol du plus petit détail perceptible par le satellite tandis que l'autre se rapporte à l'intervalle de temps au bout duquel un même point vertical à la surface pourra être de nouveau observé. Au cours de ces cinquante années d'OT, les programmes et missions spatiales se sont succédés en améliorant sans cesse les résolutions spatiales et temporelles. Ainsi, nous disposons aujourd'hui d'une archive importante de données satellitaires optiques multi-source [2], multi-échelle [3] et multi-temporelle [4]. Dans cette thèse, nous avons exploité des données provenant des satellites optiques Sentinel-2, PlanetScope, SPOT-6/7. Nous décrivons ci-après les caractéristiques de chacune de ces missions.

Sentinel-2

Sentinel-2 fait partie de la famille de satellites Sentinel de l'ESA destinée à assurer la continuité de la mission ENVISAT arrêtée en 2012. Sentinel représente le volet spatial du programme Copernicus (ex GMES) de l'Union Européenne visant à doter l'Europe d'une capacité autonome et opérationnelle en matière d'OT notamment pour la surveillance de l'environnement et la sécurité (Drusch et al., 2012). Sentinel-2 est la composante du programme devant fournir des images optiques à haute résolution spatiale mais

1. Résultat de l'interaction entre un rayonnement incident et une cible irradiée
2. Données provenant de divers satellites d'OT
3. Données disponibles à diverses résolutions spatiales de la plus grossière (kilométrique) à la plus fine (centimétrique)
4. Données disponibles sur des territoires observées de manière répétitive dans le temps

aussi temporelle permettant l'observation des surfaces émergées et la mise en place de services destinés aux situations d'urgence comme les catastrophes naturelles. La mission Sentinel-2 se réfère en fait à deux satellites identiques Sentinel-2A et Sentinel-2B lancés successivement en 2015 et 2017 et circulant en déphasage de 180° sur la même orbite héliosynchrone. Les deux satellites embarquent l'instrument MSI qui dispose de 13 bandes spectrales dans les domaines du visible et de l'infrarouge avec une résolution spatiale comprise entre 10 et 60 mètres (Tableau 1.1). Les satellites Sentinel-2 sont configurés pour fournir une image des terres émergées à intervalle de 5 jours dans le meilleur des cas. La grande richesse spectrale couplée à la capacité d'observation temporelle élevée de la mission Sentinel-2 constituent là son véritable apport dans le domaine de l'OT.

Tableau 1.1 – Caractéristiques spectrales de l'instrument MSI

Bande spectrale	Domaine spectral	Longueurs d'onde (µm)	Résolution spatiale (m)
1	Aérosols	0.433 – 0.453	60
2	Bleu	0.457 – 0.522	
3	Vert	0.542 – 0.577	10
4	Rouge	0.65 – 0.68	
5	Red Edge	0.697 – 0.712	
6	Red Edge	0.732 – 0.747	20
7	Red Edge	0.773 – 0.793	
8	Proche Infrarouge	0.836 – 0.847	10
8A	Proche Infrarouge étroit	0.855 – 0.875	20
9	Vapeur d'eau	0.935 – 0.955	60
10	Cirrus	1.365 – 1.395	
11	Infrarouge moyen	1.565 – 1.655	20
12	Infrarouge moyen	2.181 – 2.199	

PlanetScope

PlanetScope représente une constellation de nano-satellites ou CubSats de dimensions 10 × 10 × 30 centimètres exploitée par la société privée américaine Planet [5]. Cette constellation de plus de 180 nano-satellites baptisées Doves produit quotidiennement des images complètes de la Terre à très haute

5. https://www.planet.com/products/planet-imagery/

résolution spatiale variant entre 3 et 5 mètres. Elles disposent de 4 bandes spectrales dans le domaine du visible et du proche infrarouge (Tableau 1.2). La très haute résolution spatiale des images PlanetScope couplée à la répétitivité temporelle accrue des acquisitions est destinée à supporter des opérations telles que le suivi des changements climatiques, la prévision les récoltes, la gestion des catastrophes ou encore la mise au point d'applications urbaines. Les images PlanetScope sont commerciales mais peuvent être rendues librement accessibles dans la cadre de programmes d'éducation et de projets de recherche.

Tableau 1.2 – Caractéristiques spectrales des Doves PlanetScope

Bande spectrale	Domaine spectral	Longueurs d'onde (µm)
1	Bleu	0.455 – 0.515
2	Vert	0.500 – 0.590
3	Rouge	0.590 – 0.670
4	Proche Infrarouge	0.780 – 0.860

SPOT-6/7

SPOT-6 et SPOT-7 sont les deux derniers satellites issus de la famille de satellites SPOT. Lancés successivement en 2012 et 2014, cette quatrième génération de satellites vient prendre le relais du prédécesseur SPOT-5 arrivé à terme en 2015 et assurer ainsi la continuité du programme SPOT jusqu'en 2024. Équipés d'instruments identiques, les satellites SPOT-6 et SPOT-7 fournissent des images très haute résolution spatiale (jusqu'à 1.5 m) avec une capacité de revisite individuelle de 1 à 3 jours ou quotidienne en les couplant ce qui permet diverses sortes d'applications en défense et sécurité, cartographie, agriculture ou environnement. Les images SPOT-6/7 disposent d'une bande panchromatique et de bandes multispectrales dans les domaines du visible et du proche infrarouge (Tableau 1.3). Les images SPOT-6/7 sont commercialisées par la société Airbus Defence & Space [6] mais sont mises en libre accès pour les acteurs publics et institutions de recherche dans le cadre d'initiatives comme le projet EQUIPEX GEOSUD [7] ou plus récemment le dispositif DINAMIS [8].

6. https://www.airbus.com/space/earth-observation.html
7. http://ids.equipex-geosud.fr/
8. https://dinamis.data-terra.org/

Tableau 1.3 – Caractéristiques spectrales des satellites SPOT-6/7

Bande spectrale	Longueurs d'onde (µm)	Résolution spatiale (m)
Panchromatique	0.450 – 0.520	1.5
Bleu	0.450 – 0.520	6
Vert	0.530 – 0.590	
Rouge	0.625 – 0.695	
Proche Infrarouge	0.760 – 0.890	

Indices spectraux

Les indices spectraux de télédétection sont obtenus à partir d'opérations arithmétiques mises au point empiriquement sur les bandes spectrales d'images satellitaires et ont pour but de mettre en évidence des caractéristiques ou propriétés précises d'objets observés. Les indices de végétation par exemple rendent compte de l'activité photosynthétique du couvert et permettent de suivre sa dynamique. Ils sont également utilisés pour estimer sa productivité et dériver tout un panel de paramètres biophysiques comme la biomasse, l'indice de surface foliaire (LAI) ou la fraction de rayonnement photosynthétiquement actif absorbé (FAPAR). Il existe une panoplie d'indices de végétation et d'autres indices relatifs aux sols, à l'eau ou encore à l'urbain [9]. Nous détaillons dans le tableau 1.4 les indices de végétation utilisés dans le cadre de cette thèse.

9. https://www.indexdatabase.de

Tableau 1.4 – Quelques indices de végétation dérivés des bandes optiques (Abréviations : B–Bleu, G–Vert, R–Rouge, RE–Red edge, NIR–Proche infrarouge, SWIR–Infrarouge court)

Indice	Abréviation	Formule	Intérêt	Référence
Normalized Difference Vegetation Index	NDVI	$\frac{NIR-R}{NIR+R}$	Activité de la végétation en général	Rouse et al. (1973); Tucker (1979)
Normalized Difference Water Index	NDWI	$\frac{NIR-SWIR}{NIR+SWIR}$	Stress en eau	Gao (1996)
Modified Soil Ajusted Vegetation Index	MSAVI	$NIR + 0.5 - \frac{\sqrt{(2\times NIR+1)^2-8\times(NIR-R)}}{2}$	Prise en compte des effets liés au sol	Qi et al. (1994)
Enhanced Vegetation index	EVI	$2.5 \times \frac{NIR-R}{NIR+6R-7.5B+1}$	Corrige certains effets liés à l'atmosphère	Jiang et al. (2008)
Green Difference Vegetation Index	GDVI	$NIR - G$	Différence Proche infrarouge et vert	Tucker (1979)
Chlorophyll Index Green	CIgreen	$(\frac{NIR}{G})^{-1}$	Teneur en chlorophylle	Gitelson et al. (2003)
Chlorophyll Index RedEdge	CIrededge	$(\frac{NIR}{RE})^{-1}$	Teneur en chlorophylle	Gitelson et al. (2003)

1.1.2 Satellites radars

Les satellites radars d'OT aussi appelés radars imageurs ou radar à synthèse d'ouverture (RSO ou SAR) sont des systèmes de télédétection actifs. À la différence des systèmes passifs qui sont tributaires de la lumière solaire (capteurs optiques), ces derniers ont leur propre source de rayonnement incident fonctionnel de jour comme de nuit. Un SAR émet des impulsions électromagnétiques avec une fréquence ciblée dans le domaine des micro-ondes (longueur d'onde centimétrique). Les micro-ondes traversent l'atmosphère terrestre y compris ses couches nuageuses et confèrent au radar imageur l'une de ses caractéristiques les plus intéressantes à savoir l'acquisition d'images claires en toutes saisons. L'instrument SAR mesure la rétrodiffusion du signal électromagnétique c'est-à-dire la part du signal émis qui est réfléchie en direction de l'instrument. La rétrodiffusion mesurée résulte de l'action conjuguée de plusieurs facteurs. Elle est étroitement liée notamment à la nature, à l'orientation et à l'aspect de surface des objets. Par exemple, les surfaces lisses (ex. surface d'eau non agitée) auront tendance à réfléchir la totalité du rayonnement incident dans une direction symétrique qui n'est pas celle du SAR et apparaîtrons sombres. Les surfaces rugueuses par contre (ex. surface d'eau agitée) apparaîtrons plus brillantes du fait que le rayonnement incident est réfléchi dans de multiples directions. La longueur d'onde du signal SAR influence également la rétrodiffusion. Les fréquences SAR sont classées en plusieurs catégories ou bandes (voir Figure 1.1) : bande P (0.3 – 1 GHz), bande L (1 – 2 GHz), bande S (2 – 4 GHz), bande C (4 – 8 GHz), bande X (8 – 12 GHz) et bandes Ku, K, Ka (12 – 40 GHz). Les bandes de courtes fréquences et donc de grandes longueurs d'onde (comme les bandes P et L) sont plus pénétrantes que les bandes de grandes fréquences et de courtes longueurs d'onde (comme les bandes C ou X). Ces dernières sont plus sensibles à la rugosité de surface et sont réfléchies par la canopée. Néanmoins, elles permettent une meilleure perception des détails que les bandes de grandes longueurs d'onde. Il existe d'autres facteurs influençant la rétrodiffusion du signal SAR comme l'angle d'incidence du rayonnement ou le taux d'humidité des objets. Un autre aspect important des radars imageurs se rapporte à la polarisation du rayonnement électromagnétique. Les instruments SAR peuvent être configurés pour transmettre et recevoir les ondes électromagnétiques avec une polarisation verticale – V ou horizontale – H. Ainsi il existe quatre combinaisons différentes de polarisations selon la transmission et la réception des ondes : les polarisations parallèles HH et VV et les polarisations croisées HV et VH. Les polarisations ont des interactions différentes avec la surface et peuvent montrer des variations manifestes dans l'intensité du signal rétrodiffusé pour un même objet cible. Ceci contribue ainsi à fournir des

informations complémentaires sur les objets étudiés. Enfin, rappelons l'un des artefacts principaux inhérents aux images SAR : le chatoiement ou speckle. Il s'agit d'un bruit granulaire se traduisant en effet poivre et sel sur les images qui résulte de l'interférence aléatoire, à la fois constructive (responsable de tons clairs) et destructive (responsable de tons sombres), provenant de la diffusion multiple qui se produit dans chaque cellule de résolution. Cet effet est souvent atténué par des techniques de filtrage.

Tout comme dans le cas des satellites optiques, diverses missions SAR ont vu le jour au cours des dernières décennies. Celle qui a particulièrement retenu l'attention récemment en OT et également dans cette thèse, c'est la mission Sentinel-1 offrant la possibilité d'acquérir des données SAR en quasi-synchronisation avec les données optiques notamment celles de Sentinel-2. Ses caractéristiques sont décrites ci-après.

Sentinel-1

Sentinel-1 fait également partie de la famille de satellites Sentinel de l'ESA ayant vu le jour dans le cadre du programme Copernicus. C'est la composante spatiale du programme devant fournir en tout temps, de jour comme nuit, des images SAR de la surface terrestre permettant le suivi des banquises et de l'environnement arctique, la détection des glissements de terrain, la cartographie des forêts, des ressources en eau et des sols ainsi que le traitement de situations d'urgence comme les catastrophes naturelles. La mission Sentinel-1 fait également référence à deux satellites Sentinel-1A et Sentinel-1B identiques, circulant en déphasage de 180° sur la même orbite héliosynchrone. Les deux satellites sont configurés pour une période de revisite au mieux de 6 jours au nadir. Les satellites Sentinel-1 embarquent l'instrument C-SAR qui comme son nom l'indique fonctionne en bande C (5.405 GHz / 5.546 cm). Plusieurs modes d'acquisition sont fonctionnels. Sur les terres émergées, c'est le mode IW qui est le plus souvent opérationnel. Les produits distribués avec ce mode disposent de mono-polarisation (HH ou VV) ou de double polarisations (HH+HV ou VV+VH) et ont une résolution spatiale allant jusqu'à 10 mètres.

1.2 Apprentissage automatique

L'apprentissage automatique (AA) ou machine learning (ML) fait partie du vaste domaine de l'intelligence artificielle (IA). L'AA est défini comme étant l'utilisation et le développement de systèmes informatiques capables d'apprendre et de s'adapter sans suivre d'instructions explicites, grâce à des

algorithmes et modèles analysant des données et inférant des prédictions à partir des tendances représentées. Ces systèmes ont non seulement la capacité d'apprendre à partir des données mais aussi celle de s'améliorer par l'expérience, en vue d'accroître leurs performances à résoudre des tâches pour lesquelles ils ne sont pas expressément programmés. L'AA est connexe à plusieurs autres disciplines telles que les statistiques et probabilités, l'optimisation mathématique, la fouille de données ou Data Mining, la science de données ou Data science ou encore les données massives ou Big Data.

L'AA est subdivisé en plusieurs modes d'apprentissage en fonction du type de données disponibles, des algorithmes utilisés ou encore du type de problème à résoudre. Généralement, nous distinguons l'apprentissage supervisé, non supervisé et semi-supervisé.

Apprentissage supervisé : C'est probablement le mode d'apprentissage le plus commun en AA. On dispose d'un ensemble d'exemples, chaque exemple étant associé à une étiquette ou annotation ou valeur cible qu'on tente de prédire. Si cette valeur cible est discrète, on parle alors de classification ou catégorisation ou classement tandis que si elle est continue on parle de régression.

Apprentissage non supervisé : On dispose d'un ensemble de données sans aucune valeur cible associée ; le modèle doit appréhender de lui même la structure des données en déterminant des grands groupes d'appartenance non prédéfinis. On parle généralement de regroupement ou clustering.

Apprentissage semi-supervisé : C'est un mélange entre apprentissage supervisé et non supervisé dans lequel on dispose d'un petit ensemble de données annotées et d'un plus grand ensemble sans annotation. Le modèle doit tirer parti à la fois des données avec et sans valeurs cibles associées.

Dans cette thèse, nous nous intéressons essentiellement à des problèmes d'apprentissage supervisé, de classification (prédiction de l'occupation du sol) et de régression (prédiction de valeurs de rendements agricoles). L'apprentissage supervisé comporte deux phases distinctes : la phase dite d'entraînement ou d'apprentissage et celle de l'inférence ou de la prédiction. Pendant l'entraînement ou l'apprentissage, le système estime un modèle pour résoudre la tâche qui lui incombe, à partir des exemples qui lui sont fournis. En phase d'inférence, le modèle déterminé est déployé sur de nouvelles données pour prédire les valeurs souhaitées.

Le développement des algorithmes d'AA n'est pas récent. Plusieurs algorithmes rentrent aujourd'hui dans la case des approches que nous pouvons considérer comme classiques en AA. C'est le cas des forêts d'arbres décisionnels ou forêts aléatoires (Breiman, 2001) dont l'usage est très courant pour divers cas d'application de l'AA. Il s'agit d'une méthode d'apprentissage d'ensemble qui combine de multiples arbres de décisions afin d'obtenir de meilleurs performances et éviter le sur-ajustement [10]. Les arbres de décisions sont développés indépendamment sur des sous-ensembles légèrement différents des données d'entraînement en utilisant la technique de bootstrap (c'est-à-dire un échantillonnage aléatoire avec remplacement) et ensuite agrégés par vote majoritaire dans une tâche de classification ou en moyennant les prédictions dans le cas de la régression. Les forêts aléatoires sont de ce fait robustes à la multicolinéarité.

La mise en œuvre réussie des techniques d'AA traditionnelles repose avant tout sur une étape fatidique d'ingénierie de caractéristiques ou « features ». Cette étape plus ou moins lourde et/ou complexe consiste en une analyse minutieuse des données d'apprentissage afin d'en extraire les caractéristiques les plus pertinentes pour le problème posé. Des techniques de réduction de dimensionnalité de données (ex. analyse en composantes principales) sont souvent envisagées à cet effet. Néanmoins, au cours de la dernière décennie, il s'est opéré un vaste changement de paradigme qui a donné un tout nouvel élan au domaine de l'AA. Nous le devons à l'apprentissage profond (AP) ou deep learning (DL) (Lecun et al., 2015; Schmidhuber, 2015), dont l'émergence a été aussi rendue possible par la disponibilité de données d'apprentissage plus volumineux et le décuplement des capacités de calcul des ordinateurs. Contrairement aux approches classiques d'AA, les méthodes d'AP sont capables de modéliser les tâches d'AA avec un haut niveau d'abstraction des données, c'est-à-dire sans étape préalable d'ingénierie de caractéristiques. Ainsi, les méthodes d'AP sont aussi catégorisées comme des techniques d'apprentissage de représentations. Il s'agit d'un ensemble de techniques en AA permettant à un système de découvrir ou d'apprendre automatiquement les caractéristiques ou représentations nécessaires à la résolution d'une tâche spécifique à partir de données brutes. Ce type d'approche est motivé par le fait que les données du monde réel ne se prêtent pas toutes à une description commune et univoque à partir de caractéristiques manuellement extraites et que dans la plupart des cas les caractéristiques ou représentations apprises automatiquement induisent de meilleures performances.

10. Le sur-apprentissage ou sur-ajustement ou encore sur-interprétation en AA intervient quand un modèle, de par sa trop grande capacité, s'adapte ou s'ajuste trop parfaitement aux données sur lesquelles il est entraîné de telle sorte qu'il devient incapable de généraliser sur de nouvelles données non observées.

Les techniques d'AP sont aujourd'hui omniprésentes dans de nombreux domaines d'application de l'IA (ex. vision par ordinateur, traitement automatique du langage) et contribuent chaque année à des performances d'état de l'art. Toutefois, les bases ayant fait le succès de l'AP, qui repose essentiellement sur des approches par réseaux de neurones, ont été posées depuis longtemps notamment dès la proposition du perceptron (Rosenblatt, 1958) et de l'algorithme de rétropropagation du gradient ou « backpropagation » (Werbos, 1974; Rumelhart et al., 1986). Nous décrivons ci-après deux des principaux modèles d'AP sur lesquels se fondent les travaux de cette thèse, à savoir les réseaux de neurones convolutionnels et les réseaux de neurones récurrents.

1.2.1 Réseaux de neurones convolutionnels

Les réseaux de neurones convolutionnels ou convolutifs (CNNs ou ConvNets) représentent une catégorie de réseaux de neurones artificiels dits à propagation avant ou acycliques. Historiquement, leur développement est inspiré par le perceptron multi-couche (MLP) à rétropropagation (Rumelhart et al., 1986) qui est un exemple phare de réseaux de neurones acycliques. Un MLP (Figure 1.2) est composé d'une couche d'entrée, d'une ou plusieurs couches intermédiaires dites cachées et d'une couche de sortie. Chaque couche comporte un certain nombre d'unités ou neurones qui sont entièrement connectées aux neurones des autres couches.

Bien que l'incapacité initiale du MLP à résoudre des problèmes non linéaires (exemple de la fonction XOR (Minsky and Papert, 1969)) ait été solutionné grâce à l'algorithme de rétropropagation, la croissance exponentielle du nombre de paramètres lié aux multiples connexions entre neurones des différentes couches représente un inconvénient majeur qui le rend sujet au sur-apprentissage, notamment sur des données de type image. Ainsi, LeCun et al. (1989), à qui est communément attribué la paternité des CNNs, ont proposé le réseau LeNet-5, appliqué à la reconnaissance de chiffres de codes postaux manuscrits aux États-Unis, dans lequel les trois des cinq couches du réseau emploient des convolutions (Figure 1.3). Trois idées importantes au moins sous-tendent l'utilisation des convolutions (Goodfellow et al., 2016) :

- les couches convolutives sont parcimonieuses car elles limitent le nombre de connexions ou neurones d'un même champ réceptif [11] occupant par la même occasion moins de mémoire contrairement aux couches entièrement connectées ; on parle d'interaction ou de connectivité éparse
- les paramètres de convolution sont partagés ou identiques entre les

11. Le champ réceptif dans un CNN est la portion de l'espace d'entrée qui est associée à une et même convolution.

champs réceptifs
- ainsi, le résultat de la convolution est équivariant aux translations de l'image en entrée.

En général, un CNN est aujourd'hui défini par l'empilement de plusieurs blocs successifs d'extraction de caractéristiques et d'un bloc de sortie. Les blocs d'extraction de caractéristiques sont composés de couches de convolution activées par une fonction non linéaire et éventuellement de couches de pooling. Le bloc de sortie est généralement constitué de couches entièrement connectées dont la dernière est associée à une perte ou fonction de coût. Les différents blocs peuvent également faire intervenir des mécanismes de régularisation. Nous détaillons ci-après les principales notions associées à ces composantes des CNNs.

Couches de Convolution

Les couches de convolution, comme leur nom l'indique, reposent sur des opérations mathématiques appelées convolutions. L'opération de convolution est très utilisée en traitement de signal. Soient $f(t)$ et $g(t)$, deux signaux 1D, leur convolution discrète s'écrit :

$$(f * g)(t) = \sum_{k=-\infty}^{+\infty} f(k)g(t-k) \tag{1.1}$$

Dans le vocabulaire des CNNs, le premier terme de la convolution désigne le signal d'entrée tandis que le second terme représente le noyau ou filtre convolutif qui est optimisé par rétropropagation du gradient. La notion de convolution est également extensible à plus d'une dimension à la fois (ex. 2D ou 3D). Ainsi, les CNNs peuvent traiter des signaux 1D (séries temporelles), des images ainsi que des volumes (vidéos). En deux dimensions, la convolution 2D entre une image I et un noyau convolutif K s'écrit :

$$(I * K)(i,j) = \sum_{m} \sum_{n} I(i-m, j-n)K(m,n) \tag{1.2}$$

avec (i,j) les indices en lignes et colonnes de l'image I et (m,n) ceux du noyau convolutif. Notons néanmoins que pour des raisons pratiques sans incidence sur le fonctionnement des CNNs, l'implémentation de l'opération de convolution adoptée par les bibliothèques de ML s'apparente plutôt à la corrélation croisée, en ce sens où les indices de l'image et du noyau convolutif

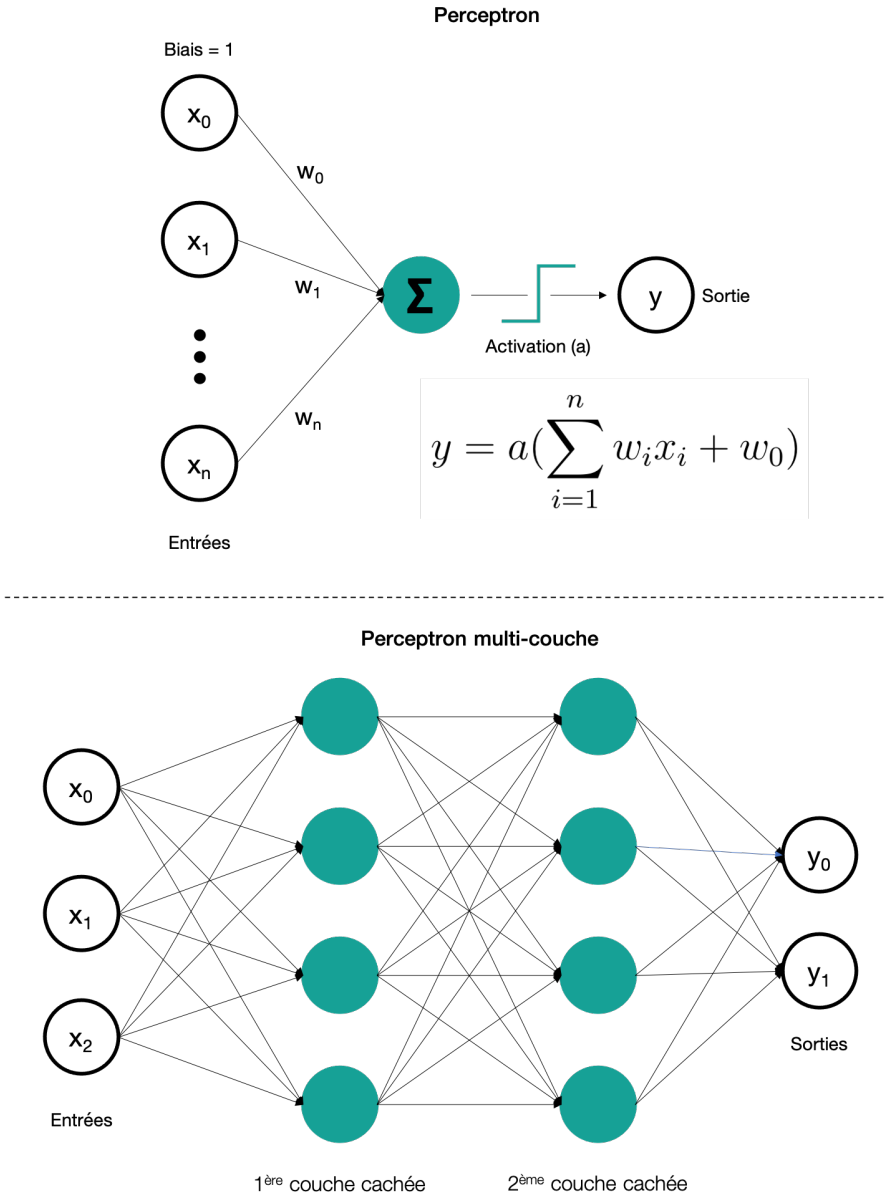


Figure 1.2 – Aperçus d'un perceptron et d'un perceptron multi-couche à deux couches cachées

sont parcourus dans le même sens. Ceci est illustré par la figure 1.4 et se traduit par l'équation suivante :

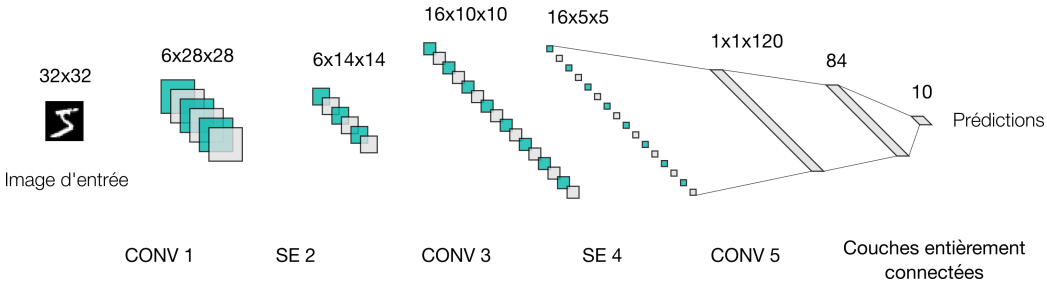


Figure 1.3 – Représentation de l'architecture Lenet-5. Les opérations de convolution CONV 1 et CONV 3 sont suivies par de sous-échantillonnages (SE 2 et SE 4).

$$(I * K)(i,j) = \sum_m \sum_n I(i+m, j+n) K(m,n) \tag{1.3}$$

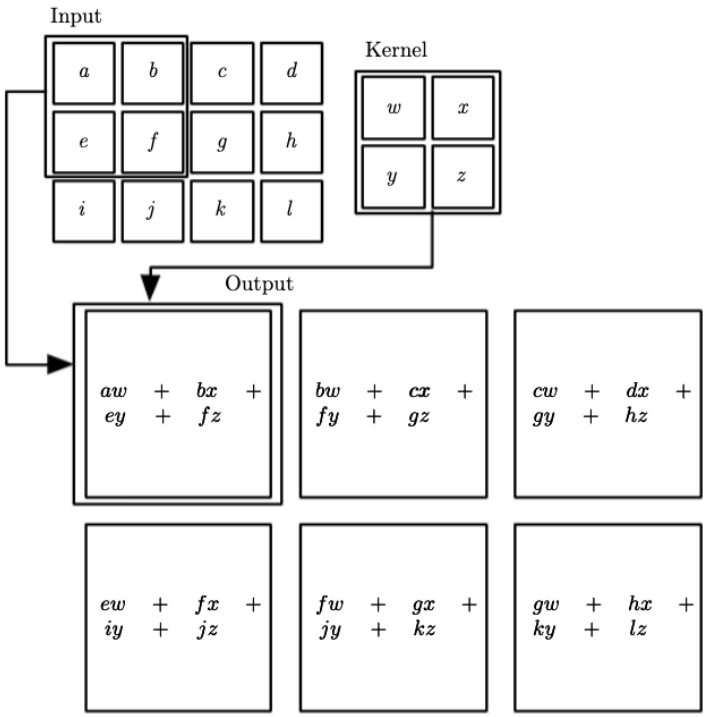


Figure 1.4 – Exemple d'une convolution 2D pour un CNN (extrait de Goodfellow et al. (2016), p. 325)

Le résultat des opérations de convolution est ensuite activé par une fonction non linéaire, ce qui produit une carte d'activation ou carte de caractéristiques. En général, une couche de convolution implique l'utilisation simultanée de plusieurs filtres convolutifs ce qui donne lieu à un ensemble de cartes d'activation en sortie. Ainsi, la profondeur du volume de sortie est égale au nombre de noyaux de convolution utilisés. La taille du volume de sortie dans les autres dimensions (d) est déterminée par la formule générique suivante :

$$d = \frac{i - k + 2 \times p}{s} + 1 \tag{1.4}$$

où i est la taille en entrée ; k est la taille du filtre de convolution généralement identique dans toutes les dimensions ; p ou « padding » correspond à une marge remplie de 0 à la frontière du volume d'entrée ; si sa valeur est égale à $\frac{k-1}{2}$, la taille de sortie est identique à celle de l'entrée ; s ou « strides » correspond à un pas de chevauchement du noyau de convolution sur l'entrée ; plus est grand, plus la taille de sortie est réduite.
L'opération de convolution de base telle que présentée précedemment (Équations (1.2) et (1.3)) est réalisée avec un pas $s = 1$ et une marge $p = 0$. Il existe encore d'autres variantes de la convolution de base qui ne sont pas abordées dans le cadre de cette thèse comme les convolutions dilatées ou à trous et les convolutions transposées.

Fonctions d'activation

Les fonctions d'activation sont des transformations mathématiques appliquées au signal en sortie d'un neurone artificiel. Le terme « activation » provient du potentiel d'action en réponse à la stimulation d'un neurone biologique. Les fonctions d'activation sont non linéaires, dérivables et peuvent être bornées ou saturantes comme la fonction sigmoïde (σ), la tangente hyperbolique ($tanh$) ou non saturantes comme la fonction Unité Linéaire Rectifiée (ReLU) et ses variantes (ex.ELU ou Leaky ReLU) et récemment la fonction GELU (Figure 1.5). En pratique, la fonction ReLU, définie comme $f(x) = max(0, x)$, est la plus populaire actuellement pour l'entraînement des modèles d'AP, y compris pour les CNNs. Elle permet un apprentissage plus rapide des réseaux de neurones par rapport aux fonctions d'activation historiques comme $tanh$ ou σ (Glorot et al., 2011).

Couches de Pooling

Dans un CNN, une couche de pooling peut être appliquée à la sortie d'une couche de convolution c'est-à-dire aux cartes d'activation. Il s'agit d'une opé-

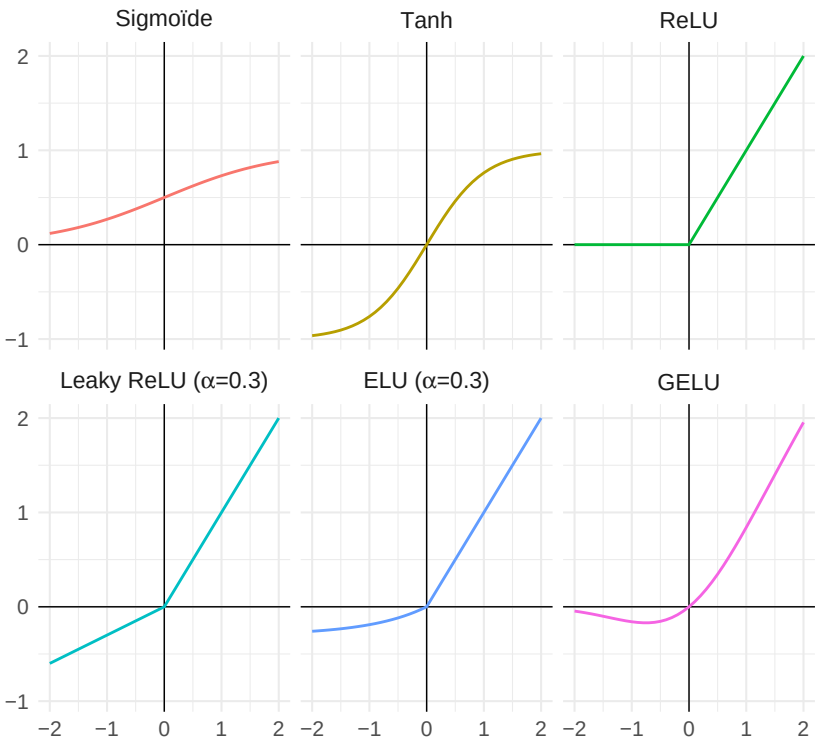


Figure 1.5 – Quelques exemples de fonctions d'activation

ration destinée à agréger et à préserver les caractéristiques ou représentations extraites de sorte à les rendre invariantes à de légères translations des données en entrée. L'opération la plus courante est celle du Max Pooling (Zhou and Chellappa, 1988) consistant à prendre successivement la valeur maximale dans un voisinage ou une fenêtre rectangulaire pouvant être mobile d'un certain pas. D'autres pratiques consistent à prendre la moyenne (Average Pooling) ou la norme L^2. La plupart du temps, l'opération de pooling s'accompagne d'un sous-échantillonnage des cartes d'activation (Figure 1.6), ce qui permet une réduction du nombre de paramètres optimisables pour la prochaine couche convolutive et de ce fait un gain en temps de calcul. Notons également que les couches de pooling ne requièrent pas de paramètres à optimiser.

Mécanismes de régularisation

Plusieurs techniques de régularisation ont été proposées afin d'améliorer la généralisation des CNNs et communément celle des modèles d'AP sur de nouvelles données. Elles peuvent consister en l'occurrence à introduire des

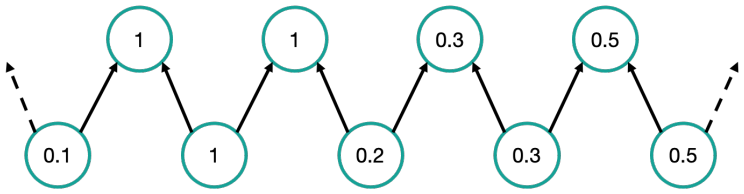


Figure 1.6 – Opération de Max Pooling accompagnée d'un sous échantillonnage avec une taille de voisinage de 2 et un pas de 1 (adapté de Goodfellow et al. (2016))

pénalités de type norme L^1 ou L^2 sur la fonction de coût en phase d'apprentissage. Une autre technique probante est celle du Dropout (Hinton et al., 2012; Srivastava et al., 2014) consistant à supprimer aléatoirement en phase d'entraînement, les connexions entre neurones avec un certain taux. Il existe également la normalisation par lot ou « batch normalization » (Ioffe and Szegedy, 2015) qui consiste à standardiser la distribution des cartes d'activation. Cette technique a également pour intérêt d'accélérer l'entraînement des réseaux convolutifs profonds.

1.2.2 Réseaux de neurones récurrents

Les réseaux de neurones récurrents (RNNs) font partie de la catégorie des réseaux de neurones artificiels dit cycliques ou à boucles. Les RNNs sont spécialisés pour le traitement de données séquentielles (ex. texte, séries temporelles) et offrent une flexibilité à la taille des séquences d'entrée qui peut être variable d'un exemple de séquence à l'autre. La récurrence dans les RNNs véhicule notamment l'idée de « mémoire », puisque les éléments antérieurs de la séquence déterminent l'état interne du réseau qui est utilisé pour calculer sa sortie au pas suivant de la séquence. Enfin, les RNNs ont également la propriété de partage des paramètres au travers des différents pas de la séquence, ce qui leur permet de garder un nombre constant de paramètres à optimiser, quelle que soit la longueur des séquences considérées.

Le RNN traditionnel ou simple RNN aussi appelé RNN « vanille » est illustré à la figure 1.7. Son fonctionnement diffère de celui d'un MLP à une couche cachée à quelques exceptions près. Soient T la taille d'une séquence x traitée par un simple RNN de I neurones ou unités d'entrée, H unités cachées et K unités de sortie, les valeurs y_h^t de l'unité cachée h, $h \in [1, H]$ et y_k^t de l'unité de sortie k, $k \in [1, K]$ au pas t de la séquence sont obtenues par les équations suivantes :

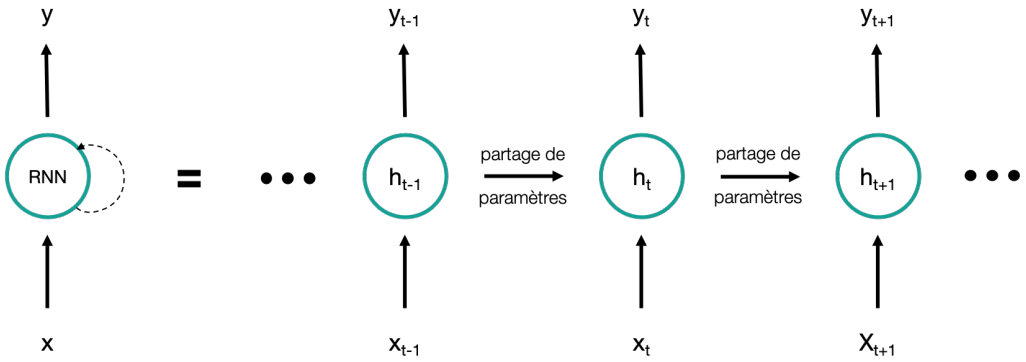


Figure 1.7 – Représentation d'un RNN vanille

$$y_h^t = \theta_h \left(\sum_{i=1}^{I} w_{ih} x_i^t + \sum_{h'=1}^{H} w_{h'h} y_{h'}^{t-1} \right) \tag{1.5}$$

$$y_k^t = \theta_k \left(\sum_{k=1}^{K} w_{hk} y_h^t \right) \tag{1.6}$$

où x_*^t est la valeur d'entrée au pas t de la séquence et θ_* une fonction d'activation, généralement une *tanh* ou fonction σ.

En fonction de la tâche de modélisation séquentielle envisagée, diverses configurations peuvent être adoptées pour les entrées et sorties du RNN (Figure 1.8). Dans cette thèse, nous nous intéressons en particulier aux architectures à plusieurs entrées et une ou plusieurs sorties. Les entrées seront en général équivalentes aux séquences temporelles tandis que, soit la sortie du RNN au dernier pas de la séquence sera directement utilisée ou soit ses sorties aux différents pas de la séquence seront combinées pour la tâche de classification ou de régression.

Les paramètres des RNNs sont optimisés différemment de ceux des réseaux de neurones à propagation avant, en utilisant d'autres algorithmes d'apprentissage comme l'algorithme de rétropropagation du gradient à travers le temps (Williams and Zipser, 1995). L'apprentissage des RNNs simples souffre néanmoins de problèmes majeurs de disparition du gradient (gradient évanescent) et d'explosion du gradient. De plus, les RNNs simples de par leur nature peuvent être adaptés à la modélisation des dépendances à court et moyen terme mais sont très vite limités dans la modélisation de dépendances à long terme. Ainsi, des modifications ont été apportées aux RNNs

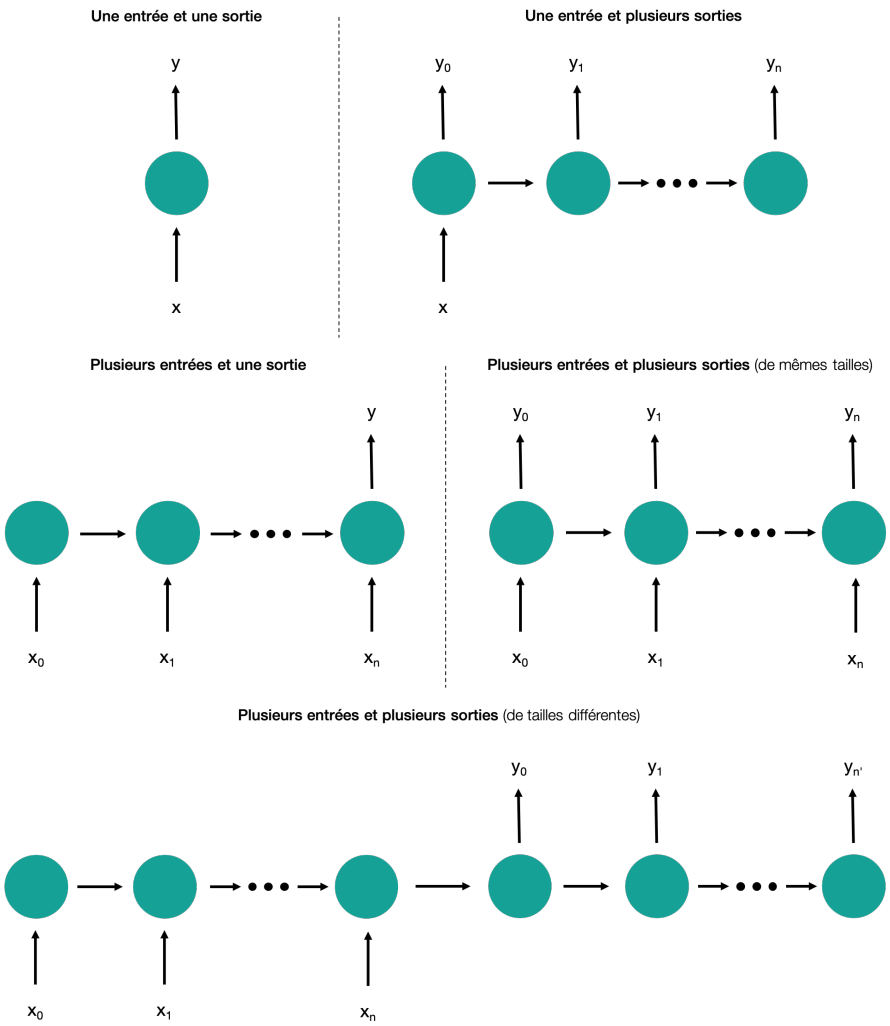


Figure 1.8 – Diverses configurations de RNN en fonction des entrées et des sorties

simples afin de pallier leurs inconvénients majeurs. Nous décrivons ci-après deux architectures plus efficaces de RNN qui avec l'avènement de l'AP, ont connu des succès retentissants dans les domaines du traitement automatique du langage (traduction automatique, analyse de sentiments) ou de la reconnaissance vocale.

LSTM

La cellule LSTM (Long-Short Term Memory) (Hochreiter and Schmidhuber, 1996; Gers, 2001) forme l'une des architectures de RNN les plus popu-

laires. Le LSTM vient pallier la disparition et l'explosion du gradient ainsi que la modélisation des dépendances à long terme, en introduisant expressément une mémoire ou état interne dans la cellule de même que trois portes (entrée, oubli et sortie) contrôlant le flux d'information qui y est accumulé. Les différentes portes présentent des valeurs entre 0 (porte fermée) et 1 (porte ouverte) en raison de leur activation par une fonction sigmoïde. La porte d'entrée détermine le degré d'ajout de nouvelles informations à la mémoire existante tandis que la porte d'oubli module le degré d'oubli de la mémoire existante. Les portes d'entrée et d'oubli déterminent à elles deux le nouvel état interne de la cellule LSTM. La porte de sortie quant à elle module le contenu de la nouvelle mémoire interne qui est exposé en sortie de la cellule pour la prochaine étape de la récurrence. Le fonctionnement d'une cellule LSTM est illustré par la Figure 1.9.

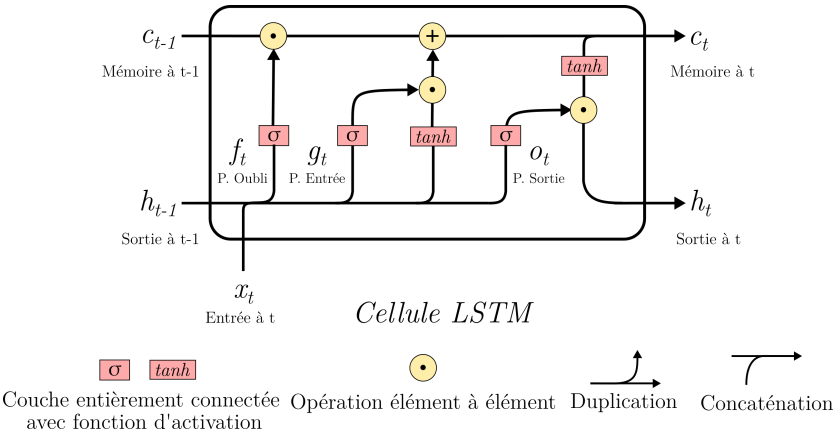


Figure 1.9 – Illustration du fonctionnement d'une cellule LSTM à une unité

Les équations suivantes régissent respectivement les portes d'entrée (g^t), d'oubli (f^t), de sortie (o^t) ainsi que l'état interne (c^t) et la sortie (h^t) :

$$g_j^t = \sigma\left(b_j^g + \sum_i W_{ji}^g x_i^t + \sum_i U_{ji}^g h_i^{t-1}\right) \tag{1.7}$$

$$f_j^t = \sigma\left(b_j^f + \sum_i W_{ji}^f x_i^t + \sum_i U_{ji}^f h_i^{t-1}\right) \tag{1.8}$$

$$o_j^t = \sigma\left(b_j^o + \sum_i W_{ji}^o x_i^t + \sum_i U_{ji}^o h_i^{t-1}\right) \tag{1.9}$$

$$c_j^t = f_j^t c_j^{t-1} + g_j^t tanh\left(b_j^c + \sum_i W_{ji}^c x_i^t + \sum_i U_{ji}^c h_i^{t-1}\right) \tag{1.10}$$

$$h_j^t = tanh(c_j^t) o_j^t \tag{1.11}$$

GRU

La cellule GRU (Gated Recurrent Unit) (Cho et al., 2014; Chung et al., 2014) est une simplification de la cellule LSTM. Contrairement au LSTM, la cellule GRU ne présente plus de mémoire interne. Elle fusionne également les portes d'entrée et d'oubli du LSTM en une nouvelle porte de mise à jour et contrôle ainsi simultanément les degrés d'oubli et de mise à jour de la cellule. La porte de sortie est pour sa part transformée en une porte de réinitialisation. Celle-ci contrôle la part de la sortie précédente qui intervient dans le calcul de la nouvelle sortie de la cellule. L'ensemble de ces modifications permettent à la cellule GRU d'atteindre des performances similaires au LSTM en ayant moins de paramètres à optimiser et un coût en temps de calcul plus réduit. Le fonctionnement d'une cellule GRU est illustré par la Figure 1.10.

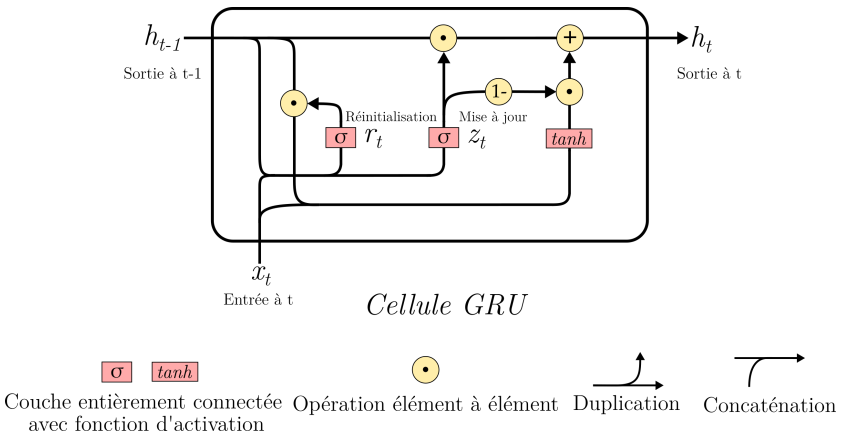


Figure 1.10 – Illustration du fonctionnement d'une cellule GRU à une unité

Les équations suivantes régissent respectivement les portes de mise à jour (z^t) et de réinitialisation (r^t) et la sortie de la cellule GRU (h^t) :

$$z_j^t = \sigma\left(b_i^z + \sum_i W_{ji}^z x_i^t + \sum_i U_{ji}^z h_i^{t-1}\right) \tag{1.12}$$

$$r_j^t = \sigma\left(b_i^r + \sum_i W_{ji}^r x_i^t + \sum_i U_{ji}^r h_i^{t-1}\right) \tag{1.13}$$

$$h_j^t = z_j^t h_j^{t-1} + (1 - z_j^t) tanh\left(b_i^h + \sum_i W_{ji}^h x_i^t + \sum_i U_{ji}^h (r_i^t h_i^{t-1})\right) \tag{1.14}$$

Mécanismes d'attention

Dans le domaine de la traduction automatique de texte, que ce soit avec un réseau LSTM ou GRU, les architectures de RNN adoptées sont pour la plupart composées d'un encodeur et d'un décodeur (Sutskever et al., 2014). Ce type d'architecture que l'on désigne sous le nom de modèle « $Seq2Seq$ » s'apparente à un cas à plusieurs entrées et sorties de tailles différentes (voir Figure 1.8). Un modèle $Seq2Seq$ transforme une séquence d'entrée en une autre séquence de sortie de taille variable. L'encodeur compresse ou encapsule la séquence d'entrée en un état caché (sortie de l'encodeur au dernier pas de la séquence d'entrée) qui est ensuite traité par le décodeur pour générer la séquence de sortie. Cette approche est néanmoins limitée par la taille de la séquence d'entrée. Plus celle ci est longue, plus son encapsulation dans l'état caché n'est plus efficace et plus les modèles $Seq2Seq$ ont du mal à être performant dans la tâche de traduction. Cette limitation a ainsi permis à l'émergence des mécanismes d'attention (MA).

L'attention a été originellement introduite par Bahdanau et al. (2015) (Figure 1.11). Au lieu que le décodeur traite uniquement l'état caché de l'encodeur au dernier pas de la séquence d'entrée, les différentes sorties intermédiaires sont pondérées selon leur importance relative pour produire un vecteur de contexte. Ce dernier est utilisé par le décodeur pour générer la séquence de sortie. Ainsi, un MA permet à un modèle de se focaliser sur les parties les plus importantes de la séquence d'entrée aux différentes étapes de génération de la séquence de sortie. Dans le processus d'attention initié par Bahdanau et al. (2015), des scores d'alignement sont tout d'abord calculés en tenant respectivement compte des différentes sorties intermédiaires de l'encodeur (h_s) et de la dernière sortie du décodeur précédant la génération de la nouvelle sortie (h_t). Les scores d'alignement reflètent ou quantifient le degré d'attention que le décodeur accordera à chacune des sorties de l'encodeur lors de la production de la sortie suivante. Ils sont calculés comme suit :

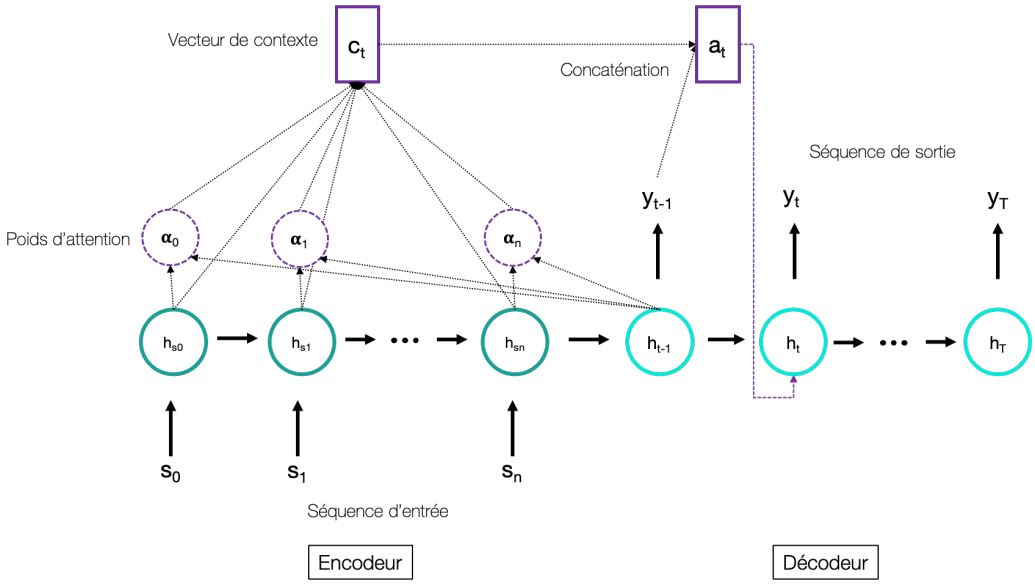


Figure 1.11 – Représentation du mécanisme d'attention classique (figure adaptée de (Luong et al., 2015))

$$e_{ts} = tanh(Wh_t + Uh_s) \tag{1.15}$$

Une autre façon d'obtenir les scores d'alignement réside également dans la multiplication (Luong et al., 2015) :

$$e_{ts} = h_t W h_s \tag{1.16}$$

Les poids d'attention (α_{ts}) sont alors calculés à partir des scores d'alignement en appliquant une fonction $SoftMax$. Cette dernière assure que les poids d'attention se somment à 1 tel des valeurs de probabilités.

$$\alpha_{ts} = \frac{exp(e_{ts})}{\sum_{s'} exp(e_{ts'})} \tag{1.17}$$

Le vecteur de contexte (c_t) par la suite obtenu est une somme pondérée des poids d'attention et des différentes sorties intermédiaires de l'encodeur.

$$c_t = \sum_{s'} \alpha_{ts'} h_{s'} \tag{1.18}$$

Enfin, le décodeur procède à la génération de la nouvelle sortie, à partir de la concaténation du vecteur de contexte et de la sortie précédente qui lui est passée en entrée.

$$a_t = tanh(W^c[c_t, h_t]) \tag{1.19}$$

En plus de permettre aux modèles de se focaliser sur ce qui est important dans le signal d'entrée, les MA et particulièrement les poids d'attention leur apportent aussi un aspect explicatif (Figure 1.12). En effet, puisque ces derniers forment une combinaison convexe, ils peuvent être analysés pour interpréter de manière probabiliste les décisions des modèles d'AP souvent qualifiés de « boîtes noires ». Ainsi, les MA contribuent d'une part à l'IA « explicable ». Le MA initial a été par la suite approfondi par d'autres études notamment celle de Vaswani et al. (2017) sur les « Transformers ». Il s'agit de modèles purement basé sur l'attention à la différence d'autres approches qui mêlent CNN ou RNN et MA (Xu et al., 2015; Britz et al., 2017). Les Transformers et modèles dérivés (ex. BERT (Devlin et al., 2019)) ont également été à la base de performances d'état de l'art ces dernières années en traitement automatique du langage.

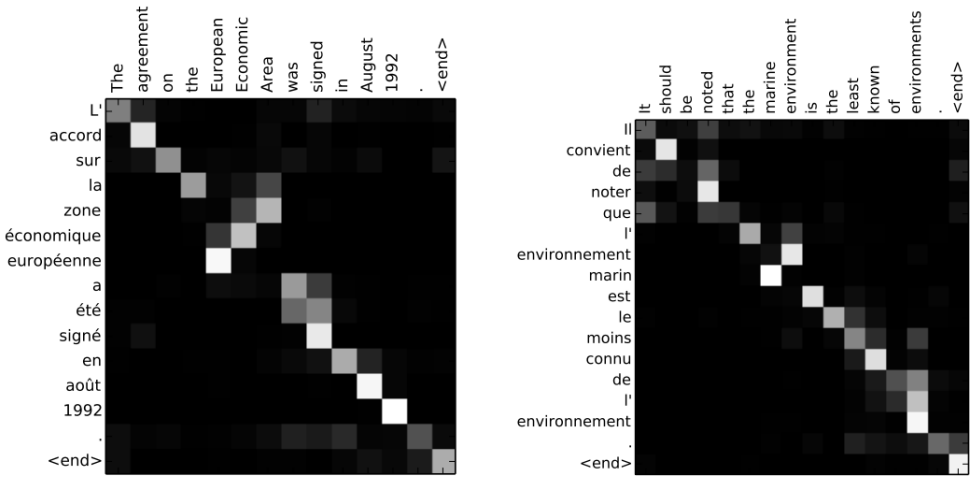


Figure 1.12 – Exemples d'analyse des poids d'attention sur la traduction de texte de l'anglais vers le français (extrait de Bahdanau et al. (2015), p. 6). Chaque carré au sein des matrices représente le poids accordé à un mot source (en anglais) pour la traduction d'un mot cible spécifique (en français). Les poids sont représentés en niveau de gris (du noir vers le blanc pour des poids entre 0 et 1).

1.2.3 Métriques pour l'évaluation des modèles d'apprentissage supervisé

Nous détaillons ci-après les métriques utilisées dans le cadre de cette thèse pour l'évaluation de nos modèles. Nous distinguons les métriques pour évaluer la classification de l'occupation du sol (Précision gobale, F1-score et cœfficient Kappa) et la prédiction des valeurs rendements agricoles (cœfficient de détermination, racine de l'écart quadratique moyen et erreur absolue moyenne)

Métriques pour la classification

Précision globale La précision ou exactitude globale représente la proportion des prédictions correctes effectuées par un modèle. Elle varie entre 0 et 1 et est donnée par l'équation suivante :

$$Precision\ globale = \frac{1}{n}\sum_{i=1}^{n}(y_i\ =\ \hat{y}_i) \tag{1.20}$$

où n est le nombre total d'observations ; y_i et $\hat{y}_i$ représentent respectivement la valeur vraie de classe et la prédiction du modèle pour l'observation i ; $(y_i\ =\ \hat{y}_i)$ équivaut à 1 si la prédiction du modèle est correcte et 0 sinon.

F1-score Le F1-score ou F-mesure ou cœfficient de Sørensen–Dice est une moyenne harmonique entre la précision et le rappel d'un classifieur. La précision ($\frac{VP}{VP+FP}$) dans le cas d'une classification binaire correspond à la part de vrai positifs (VP), c'est-à-dire au nombre de prédictions effectives de cette classe, sur le nombre total de prédictions positives effectuées par le modèle comprenant les faux positifs (FP). Le rappel ou la sensibilité est dans ce cas la part de vrai positifs sur le nombre total d'observations effectivement positives, comprenant les faux négatifs (FN) c'est-à-dire les prédictions manquées par le classifieur. Le F1-score est particulièrement utile pour évaluer les performances de modèles sur des tâches de classification déséquilibrées où une ou plusieurs classes sont sur-représentées ou l'inverse en matière d'observations par rapport aux autres classes (ex. en classification de l'occupation du sol), car il ne prend pas en compte les vrais négatifs (VN) à l'instar de la précision globale que nous pourrions réécrire : $\frac{VP+FN}{VP+FP+VN+FN}$.

Ceci permet de mettre en exergue le comportement réel des modèles en cas de classe sous-représentée (nombre d'observations souvent largement inférieur à l'ensemble des vrais négatifs), ce qui est occulté par la précision

globale. Dans un contexte de classification binaire, le F1-score est donné par l'équation (1.21).

$$F1 = \frac{2}{precision^{-1} + rappel^{-1}} = 2 \times \frac{precision \times rappel}{precision + rappel} \tag{1.21}$$

Dans un contexte multi-classe comme celui que nous traitons, nous pondérons les F1-score de chaque classe par leur nombre d'échantillons respectifs (Équation (1.22)). Il varie également entre 0 et 1.

$$F1_{multi} = \frac{1}{n} \sum_{c=1}^{C} n_c \times F1_c \tag{1.22}$$

Cœfficient Kappa Le cœfficient Kappa (Cohen, 1960) s'interprète comme la proportion d'accord ou de jugements concordants entre deux évaluateurs, substitué de l'effet de l'aléatoire. Il est donné par l'équation suivante :

$$Kappa = \frac{P_o - P_e}{1 - P_e} \tag{1.23}$$

où P_o est la proportion d'accord observée entre les évaluateurs et P_e la proportion d'accord due à l'aléatoire.

$$P_o = \frac{1}{n} \sum_{i=1}^{C} n_{ii} \qquad P_e = \frac{1}{n} \sum_{i=1}^{C} \frac{n_{i.} \times n_{.i}}{n} \tag{1.24}$$

où n_{ii} est l'effectif associé à la ième ligne et colonne d'un tableau de contingence de C (nombre de classes) lignes et colonnes ; $n_{i.}$ et $n_{.i}$ sont respectivement la somme des effectifs de la ième ligne et colonne. Le cœfficient Kappa varie entre -1 (pour un accord totalement dû à l'aléatoire) et 1 (pour un accord complet).

Métriques pour la régression

Cœfficient de détermination Le cœfficient de détermination ou R^2 représente, dans une régression linéaire, la proportion de la variance dans la variable dépendante qui est expliquée par le modèle. Le R^2 varie généralement entre 0 et 1 (pour un ajustement parfait entre observations et prédictions),

mais peut être négatif quand l'ajustement est très mauvais. Il est donné par l'équation suivante :

$$R^2 = 1 - \frac{\sum_{j=1}^{n}(y_j - \hat{y_j})^2}{\sum_{j=1}^{n}(y_j - \bar{y})^2} \tag{1.25}$$

n désigne le nombre total d'observations ; y_j est la valeur observée pour le jième échantillon ; $\hat{y_j}$ est la valeur prédite pour cet échantillon et $\bar{y}$ est la moyenne des valeurs observées. Les désignations restent valables pour les équations (1.26) et (1.27).

Racine de l'écart quadratique moyen ou RMSE La RMSE donne une indication sur l'ampleur des écarts moyens de prédiction d'un modèle de régression. Comme son nom l'indique, la RMSE est calculé en prenant la racine de l'écart quadratique moyen (MSE) ou moyenne du carré des écarts entre observations et prédictions. La MSE et de ce fait la RMSE pénalisent fortement les grandes erreurs de prédictions en surestimation comme en sous-estimation. La prise de la racine a pour but d'exprimer la valeur du RMSE dans la même échelle de mesure que la variable dépendante. La RMSE est sans borne supérieure. Plus il est proche de 0, meilleures sont les prédictions du modèle évalué. Il est calculé de la manière suivante :

$$RMSE = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_j - \hat{y_j})^2} \tag{1.26}$$

Erreur moyenne absolue ou MAE Tout comme la RMSE, la MAE mesure l'écart moyen des prédictions d'un modèle de régression. À la différence du RMSE, la MAE considère plutôt la valeur absolue des erreurs de prédictions. La MAE est également sans borne supérieure et plus il est proche de 0, meilleures sont les prédictions. Il reste par ailleurs inférieur ou égal en grandeur par rapport au RMSE. Leurs valeurs sont égales si l'amplitude des écarts est la même pour tous les échantillons. La MAE se calcule de la manière suivante :

$$MAE = \frac{1}{n}\sum_{j=1}^{n}|y_j - \hat{y_j}| \tag{1.27}$$

Chapitre 2

Sites d'étude et Données utilisées

Sommaire

2.1 Sites d'étude

Les travaux de cette thèse ont été menés sur divers sites d'étude aux caractéristiques paysagères variées. Il s'agit des sites de l'île de la Réunion, de la Dordogne et du bassin arachidier du Sénégal. Sur les différentes zones d'étude, des données relatives à la couverture au sol et/ou aux rendements des parcelles agricoles ont été collectées sur le terrain.

2.1.1 Site de l'île de la Réunion

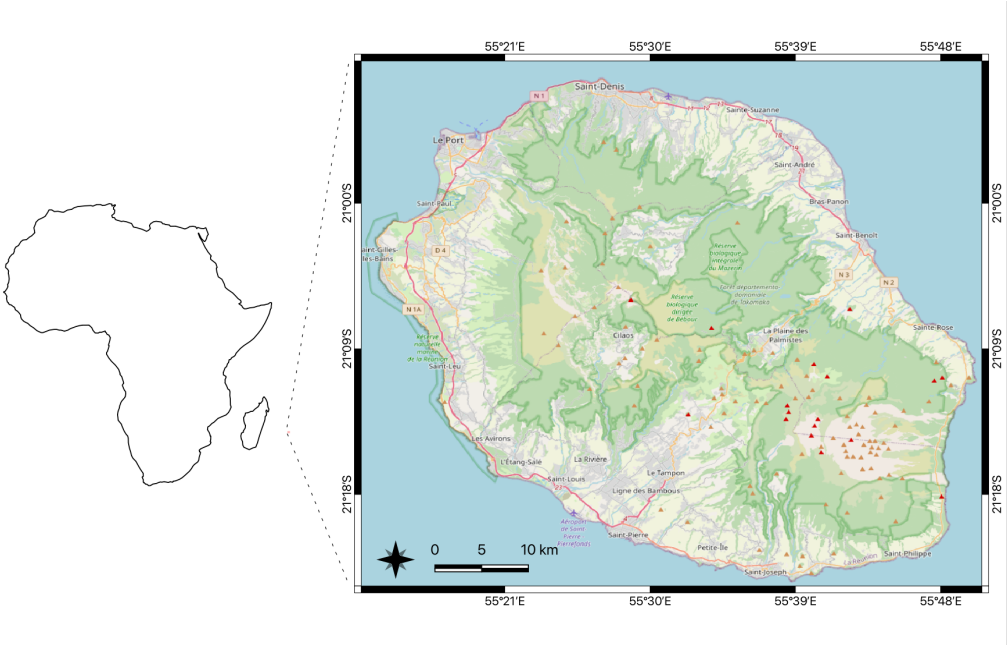


Figure 2.1 – Localisation de l'île de la Réunion (le fond de carte utilisé est issu de OpenStreetMap)

L'île de la Réunion est un département français d'outre-mer situé dans l'hémisphère sud, plus précisément à l'est de l'Afrique près de Madagascar dans l'océan indien (Figure 2.1). La Réunion s'étend environ sur 2500 km^2 et est connue pour son paysage très accidenté ainsi que ses volcans que sont le piton des Neiges à l'ouest (éteint) et le piton de la Fournaise à l'est (actif). Le point culminant de l'île se trouve sur le piton des Neiges à 3071 m d'altitude. En plus de ses volcans, la Réunion abrite en son centre les cirques naturels de Mafate, Salazie et Cilaos qui font partie intégrante du parc national de la Réunion, inscrit depuis 2010 au patrimoine mondial de l'UNESCO. Bien que diversifiée dans la production de fruits et légumes, l'agriculture de l'île de

Réunion repose essentiellement sur la filière de la canne à sucre qui représente à elle seule plus de la moitié de la surface agricole utile soit un peu plus de 22 000 ha [1].

Données collectées

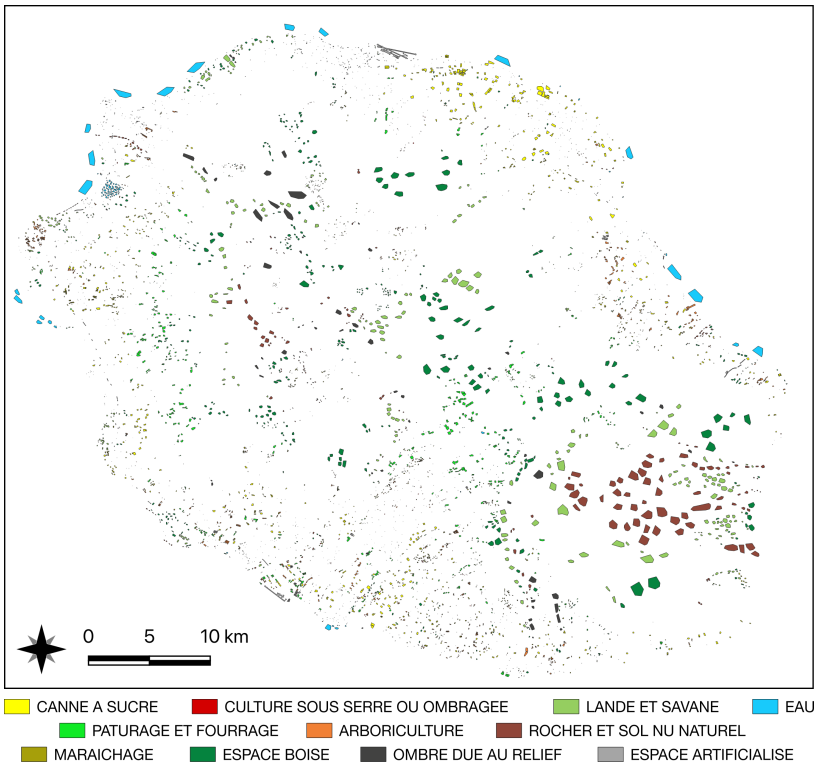


Figure 2.2 – Distribution spatiale des données collectées sur l'île de la Réunion

Les données collectées sur l'île de la Réunion sont relatives à sa couverture au sol. Il s'agit de la vérité terrain de l'année 2017 qui a été construite sur base de plusieurs sources : le RPG de l'année 2017, des relevés de points GPS issus d'une mission terrain effectuée en Juin 2017 et l'interprétation visuelle d'une image SPOT-6 à très haute résolution spatiale. Un total de 6 265 échantillons ont été collectés et sont répartis sur 11 classes d'occupation du sol (Figure 2.2 et tableau 2.1). Ceci fait de l'île de la Réunion le site le plus fourni en termes d'échantillons disponibles et le plus diversifié en matière de classes d'occupation du sol. Le jeu de données sur l'île de la Réunion est accessible publiquement sur la Dataverse du Cirad [2]. Des informations

1. Source : Chambre d'agriculture de la Réunion (https://www.reunion.chambagri.fr)
2. https://doi.org/10.18167/DVN1/TOARDN

supplémentaires sur la collecte de ces données sont reportés par Dupuy et al. (2020).

Tableau 2.1 – Nombre d'échantillons par classe sur l'île de la Réunion

Classe	Échantillons	Superficie (km^2)
1 – CANNE À SUCRE	869	8.90
2 – PÂTURAGE ET FOURRAGE	582	6.81
3 – MARAÎCHAGE	758	1.76
4 – CULTURE SOUS SERRE OU OMBRAGÉE	260	0.19
5 – ARBORICULTURE	767	3.36
6 – ESPACE BOISE	570	20.50
7 – LANDE ET SAVANE	506	15.52
8 – ROCHER ET SOL NU NATUREL	299	15.43
9 – OMBRE DUE AU RELIEF	81	5.43
10 – EAU	177	8.26
11 – ESPACE ARTIFICIALISÉ	1396	1.90

2.1.2 Site de la Dordogne

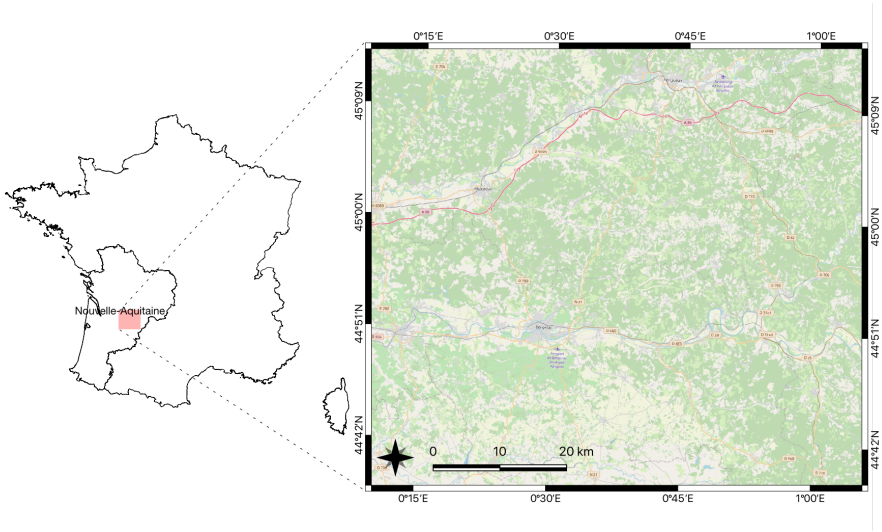


Figure 2.3 – Localisation du site de la Dordogne (le fond de carte utilisé est issu de OpenStreetMap)

Le site de la Dordogne se situe dans le département du même nom en France métropolitaine dans la région Nouvelle-Aquitaine (Figure 2.3). Le

site d'étude d'une superficie d'environ 3000 km^2 se trouve au sud-ouest du département à cheval entre les communes de Périgueux et Bergerac. Cette dernière est traversée en son sein par le fleuve Dordogne. La zone d'étude est essentiellement plate avec par endroits des collines et vallées. Le paysage est très boisé et recouvert de forêts de feuillus et de conifères ainsi que de prairies. Les terres agricoles sont pour la plupart composées d'arbres fruitiers et de vignes.

Données collectées

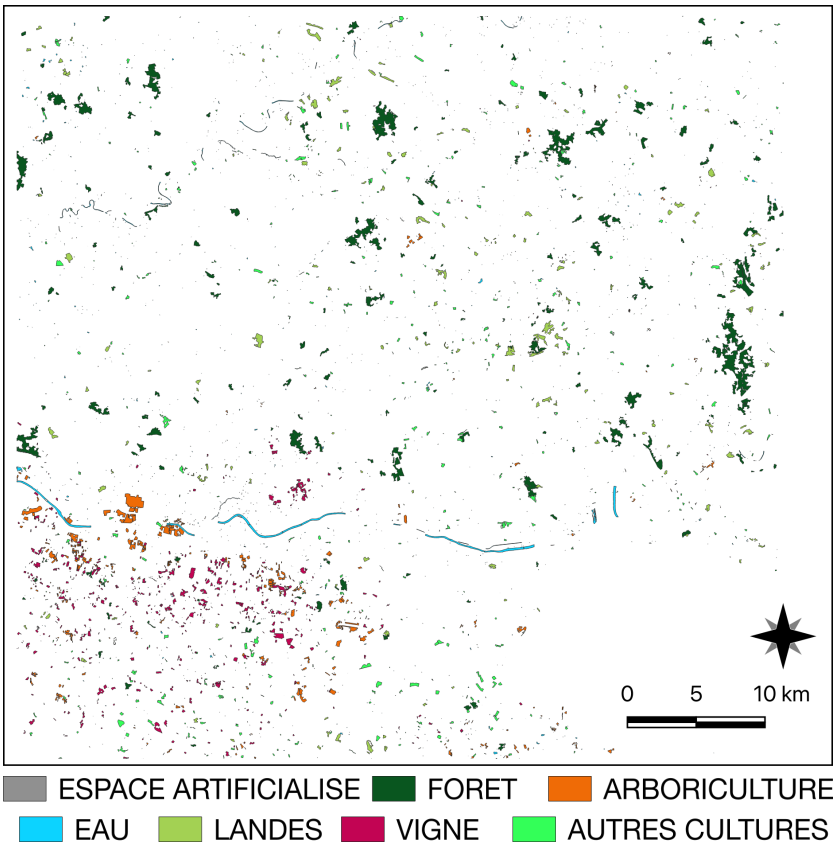


Figure 2.4 – Distribution spatiale des données collectées sur le site de la Dordogne

Les données collectées sur le site de la Dordogne concernent également sa couverture au sol. La vérité terrain est celle de 2016 et a été construite à partir du RPG de l'année 2016 ainsi que des BD TOPO et FORET de l'IGN. Un total de 3 819 échantillons ont été assemblés sur ce site et sont répartis sur 7 classes d'occupation du sol (Figure 2.4 et tableau 2.2). Ce jeu

de données n'est pas encore disponible publiquement mais peut être mis à disposition sous demande.

Tableau 2.2 – Nombre d'échantillons par classe sur le site de la Dordogne

Classe	Échantillons	Superficie (km^2)
1 – ESPACE ARTIFICIALISÉ	800	0.21
2 – EAU	800	5.06
3 – FORÊT	200	37.90
4 – LANDES	187	9.97
5 – ARBORICULTURE	632	9.74
6 – VIGNE	600	9.24
7 – AUTRES CULTURES	600	9.38

2.1.3 Site du bassin arachidier au Sénégal

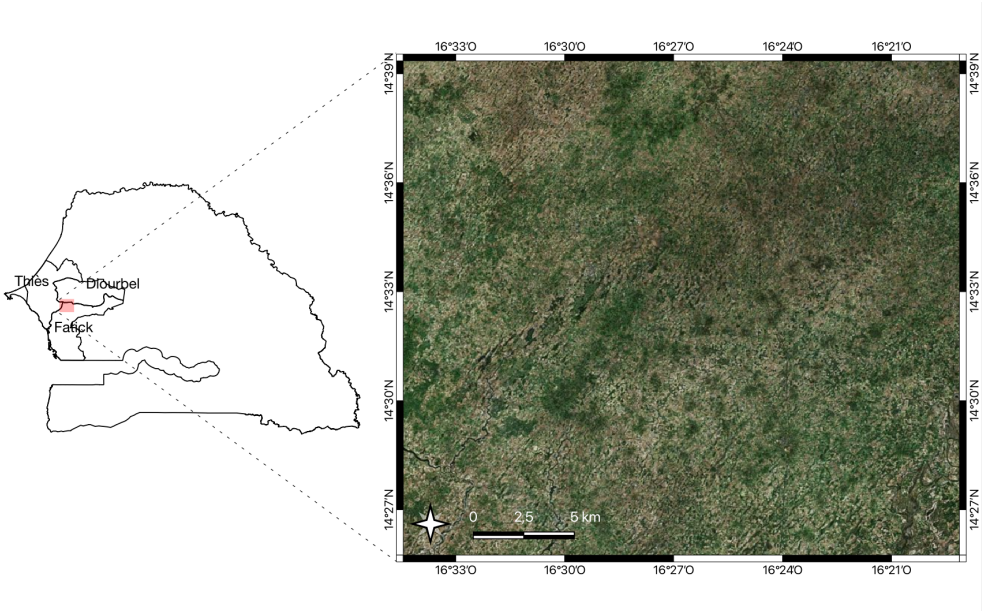


Figure 2.5 – Localisation du site du bassin arachidier au Sénégal (l'image en fond de carte provient de Bing Satellite)

Le site du bassin arachidier se situe au centre ouest du Sénégal, de part et d'autre des départements de Bambey (région de Diourbel), Fatick (région de Fatick) et M'bour (région de Thiès) (Figure 2.5). La zone d'étude est

essentiellement agricole et se situe dans une vaste plaine dont les altitudes maximales ne dépassent pas 100 m. Elle s'étend environ sur 450 km^2. On y trouve principalement des céréales notamment mil et sorgho et des légumineuses en l'occurrence l'arachide (d'où vient le nom bassin arachidier) ainsi que le niébé. Le paysage est cependant très hétérogène, caractérisé par la présence de petites parcelles (moins de 1 ha) présentant couramment des associations culturales (céréales avec légumineuses) ainsi que des arbres isolés en leur sein formant un parc arboré. Ce parc arboré est diversifié et comprend de nombreuses espèces et est dominé par le *Faidherbia albida*, légumineuse fixatrice d'azote, connu en agroforesterie pour sa phénologie inversée et ses effets stimulants sur les cultures.

Données collectées

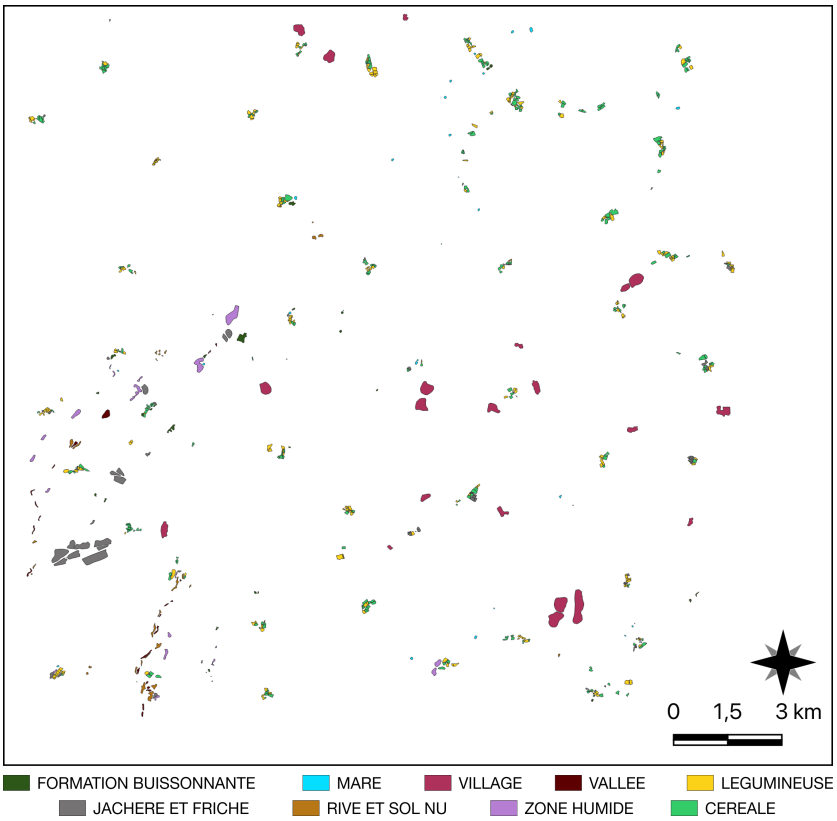


Figure 2.6 – Distribution spatiale des données d'occupation du sol collectées en 2018 sur le site du Sénégal

Les données collectées sur le site du Sénégal sont de deux types. Elles sont relatives non seulement à sa couverture au sol mais également aux rendements

des parcelles agricoles. Les données sur l'occupation du sol ont été collectées sur 2 années : celles des saisons agricoles 2018 et 2019. La vérité terrain de 2018 a été construite à partir de relevés de points GPS effectués lors d'une mission terrain en mi-saison (Juillet) et de l'interprétation visuelle d'une image PlanetScope à très haute résolution spatiale. Un total de 734 échantillons ont été collectés et sont répartis sur 8 classes (Figure 2.6 et tableau 2.3). En 2019, une nouvelle mission a consisté à recollecter la vérité terrain des échantillons dont la classe d'occupation du sol est susceptible de varier d'une année à l'autre (ex. cultures, jachères). Ainsi, un total de 669 échantillons ont été assemblés pour la vérité terrain de 2019 (Tableau 2.3). Que ce soit en 2018 ou 2019, le site du Sénégal représente la zone la moins fournie d'entre nos 3 sites en termes d'échantillons disponibles mais sans oublier que sa superficie est tout aussi moins vaste.

Tableau 2.3 – Nombre d'échantillons par classe sur le site du Sénégal

Classe	Échantillons		Superficie (km^2)	
	2018	2019	2018	2019
1 – FORMATION BUISSONNANTE	50	50	0.16	0.16
2 – JACHÈRE ET FRICHE	69	27	0.85	0.06
3 – MARE	33	33	0.07	0.07
4 – RIVE ET SOL NU	35	35	0.19	0.19
5 – VILLAGE	21	21	1.40	1.40
6 – ZONE HUMIDE	22	22	0.34	0.34
7 – VALLÉE	22	22	0.13	0.13
8 – CÉRÉALE	260	311	1.13	1.58
9 – LÉGUMINEUSE	222	148	0.93	1.01

Les rendements des parcelles agricoles ont également été collectés lors des saisons agricoles de 2018 et 2019. En plus, nous avons bénéficié de données sur les rendements de 2017, collectées dans le cadre d'autres thèses dans la zone d'étude notamment celle de Mme. Sophie DJIBA (ISRA/IRD) et de M. Adama TOUNKARA (ISRA/IRD/Cirad)[3]. Les rendements dont nous disposons ont été collectés autour du village de Diohine, au sud-ouest de la zone d'étude. La zone de collecte s'étend environ sur 17 km^2. Les parcelles cibles sont celles du mil, une culture de base dans la région qui sert directement à la consommation des ménages. La variété de mil cultivée est le *Souna* qui dispose d'un cycle court d'environ 90 jours. Les semis du mil se font généralement à sec entre la fin Mai et le début du mois de Juin en prévision du début de la saison des pluies. Le pic de l'activité chlorophyllienne intervient pendant l'hivernage (saison des pluies) au cours de la période végétative qui

3. https://ur-aida.cirad.fr/nos-recherches/theses-en-cours/tounkara

a lieu le plus souvent en mi-Août ou en début Septembre, si décalage de la saison des pluies il y a. Ensuite interviennent les phases de reproduction et de maturation des grains jusqu'à la récolte qui a lieu au plus tard en fin Octobre.

Sur les 3 saisons agricoles, un total de 66 parcelles ont été suivies : 35 en 2017, 15 en 2018 et 16 en 2019 (Figure 2.7). Sur ce total, 81% des parcelles ont un système de culture mixte essentiellement du mil associé au sorgho ou au niébé. La taille moyenne des parcelles suivies est de 0.63 ha. Dans le protocole de collecte, trois quadrats représentatifs de la parcelle de $6m^2$ sont choisis au hasard. La biomasse aérienne y est ensuite récoltée à la maturité et le rendement en grain mesuré après séchage à 70°C pendant 48 heures. D'autres informations sont par ailleurs collectées sur les parcelles suivies entre autres les dates de semis et de récolte, les dates observées des différents stades phénologiques : émergence, maturité, sénescence. Les rendements observés sont compris entre 239 kg/ha et 3278 kg/ha avec une moyenne de 1184.51 kg/ha et une médiane de 976.5 kg/ha. Ces valeurs illustrent à la fois la grande hétérogénéité spatiale et la grande variabilité inter-annuelle existante dans la zone à l'instar de 2017 et 2019 (Figure 2.8). Plusieurs facteurs peuvent en être à l'origine. Les plus communs sont le début de la saison des pluies qui détermine le calendrier agricole et donc les rendements finaux (Marteau et al., 2011), les pratiques culturales telles que les rotations ou encore la fertilité des sols et la quantité de nutriments qui y sont apportés comme l'ont récemment montré Leroux et al. (2020).

2.2 Prétraitement des données satellitaires

Comme évoqué précédemment (Section 1.1), des données provenant d'un côté de satellites optiques (Sentinel-2, SPOT et PlanetScope) et de l'autre du satellite radar Sentine-1 ont été mobilisées dans le cadre de cette thèse. Nous détaillons ci après les étapes de prétraitement spécifiques à ces données.

2.2.1 Séries temporelles Sentinel-1

Les images Sentinel-1 ont été téléchargées sur la plateforme PEPS du CNES [4]. Ce sont des produits de niveau-1 GRD, acquis en mode IW en orbites ascendante et descendante. Elles disposent de la double polarisation VV et VH. Afin d'uniformiser les données sur nos 3 zones d'étude, seules les données en orbite ascendante, orbite commune à tous les sites, ont été ultérieurement explorées. Les images ont été collectées périodiquement sous

4. https://peps.cnes.fr/

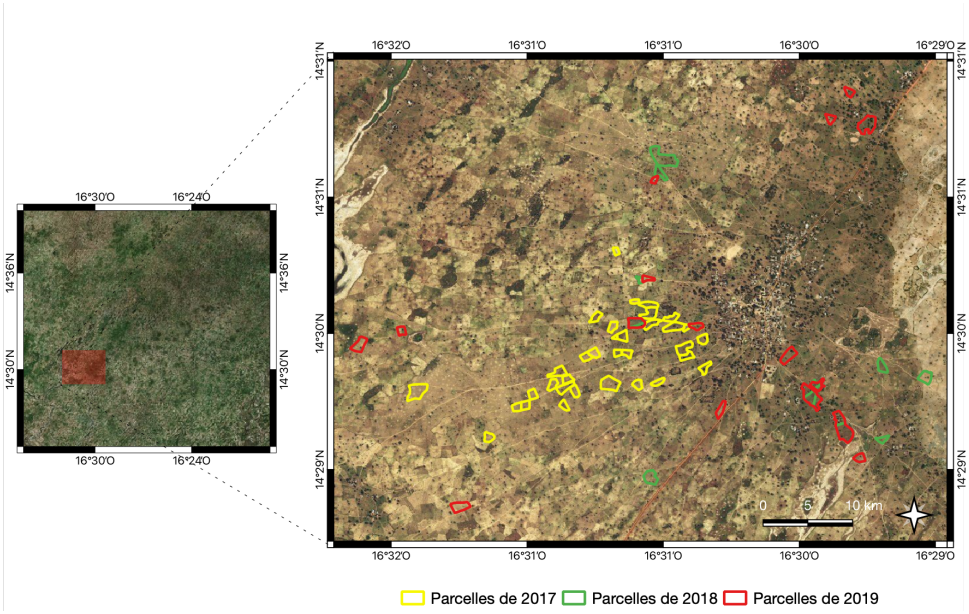


Figure 2.7 – Distribution spatiale des parcelles agricoles suivies sur les 3 années (2017, 2018 et 2019). Certaines parcelles ont été à la fois suivies en 2018 et 2019 (les images en fond proviennent de Bing Satellite).

Figure 2.8 – Répartition annuelle des rendements des parcelles

la forme de séries temporelles sur chaque site. Ces séries temporelles ont été finalement prétraitées en appliquant successivement :

- la calibration radiométrique (sigma zéro) des valeurs d'intensité du signal SAR retrodiffusé en cœfficient de rétrodiffusion
- l'orthorectification des images à la résolution spatiale de 10 mètres (tout comme les images Sentinel-2)

- et le filtrage multi-temporel (Quegan and Yu, 2001) pour débruiter le plus possible les images de la série temporelle en réduisant le chatoiement ou speckle. Ce filtre procède par combinaison linéaire des valeurs de cœfficient de rétrodiffusion SAR obtenues aux différentes dates, avec une moyenne locale estimée à partir d'une fenêtre prise autour de chaque pixel dans chaque image.

Ces étapes de prétraitement sont effectuées avec la chaîne S1Tiling [5] développée au CESBIO.

2.2.2 Séries temporelles Sentinel-2

Les images Sentinel-2 ont quant à elles été téléchargées sur l'IDS du pôle THEIA [6]. La chaîne MUSCATE du pôle THEIA traite au niveau-2A les images S2 de niveau-1C distribuées par le programme Copernicus sur une partie des surfaces continentales. Nos trois zones d'étude sont couvertes par les images traitées par le pôle THEIA. Au niveau 1C, les images sont orthorectifées et les valeurs brutes calibrées en réflectance au dessus de l'atmosphère ou Top of Atmosphere (TOA). Le niveau-2A reflète les valeurs de réflectance de surface ou Top of Canopy (TOC) après les corrections atmosphériques effectuées par la chaîne MAACS/MAJA. Les réflectances de surface fournies sont de deux types :

- le type SRE pour Surface REflectance incluant des corrections atmosphériques et d'effets d'environnement
- et le type FRE pour Flat REflectance incluant en plus des précédentes, des corrections liées aux effets de pente

C'est ce dernier que nous avons adopté afin de prendre notamment en compte le relief accidenté de l'île de la Réunion. Par ailleurs au niveau-2A, des masques de nuages, d'ombre, de surfaces d'eau et de neige sont également fournis. Les images Sentinel-2 ont également été collectées périodiquement sous la forme de séries temporelles sur chacun des sites d'étude. Étant donné la présence d'images ennuagées dans le domaine de l'optique, nous avons exclu le plus possible d'images affectées, de sorte à maintenir une vingtaine d'images collectées sur chacun des sites d'étude. Finalement, les valeurs des pixels classés comme nuages sur les masques associés aux images collectées sont linéairement interpolées pour chaque bande spectrale considérée, sur base des autres observations non manquantes de part et d'autre dans la série temporelle. Cette technique connue sous le nom de « gap-filling » (remplissage de trous) est communément adoptée dans les prétraitements des séries

5. http://tully.ups-tlse.fr/koleckt/otbsarmultitempfiltering
6. https://theia.cnes.fr

temporelles d'images satellitaires optiques (Inglada et al., 2017).

2.2.3 Images SPOT-6/7

Les images SPOT-6/7 ont été obtenues gratuitement dans le cadre du projet EQUIPEX GEOSUD. Il s'agit d'images orthorectifées et acquises sur les sites de la Réunion et de la Dordogne. Nous n'avons adopté aucun prétraitement supplémentaire pour ces images.

2.2.4 Images PlanetScope

L'accès aux images PlanetScope nous a été fourni gratuitement dans le cadre du programme d'éducation et recherche de la société Planet qui les commercialise. Les images obtenues sont de niveau-3B, orthorectifiées et acquises sur le site du Sénégal. Leur résolution spatiale est de 3 mètres. Elles ont par la suite été calibrées en réflectance TOA grâce aux valeurs de radiance et coefficients fournis dans les métadonnées des scènes.

Chapitre 3

Caractérisation de l'occupation du sol

Sommaire

3.1 Introduction

La cartographie des surfaces cultivées revêt une importance cruciale pour l'évaluation de la production agricole. À cet effet, les systèmes de suivi des cultures recourent usuellement à la synthèse de divers produits globaux d'occupation du sol présentant toutefois des incertitudes avérées quant à la localisation des surfaces cultivées. Le présent chapitre traite du premier objectif de cette thèse lié à la caractérisation de l'occupation du sol pour l'identification des surfaces cultivées à partir de méthodes d'apprentissage automatique employant des données de télédétection multi-source.

Les données de télédétection, qu'elles soient en libre accès ou non, sont mises à contribution depuis des décennies pour cartographier les dynamiques intervenant à la surface terrestre. En un premier temps dans ce chapitre, nous associons uniquement des données radar et optique publiquement accessibles c'est-à-dire des séries temporelles d'images Sentinel-1 et Sentinel-2, puis ensuite, nous évaluons l'apport d'une source optique à très haute résolution spatiale (SPOT), dont l'accès est usuellement commercial et la fréquence d'acquisition restreinte en raison des coûts d'exploitation.

Deux grands paradigmes existent lorsque des données de télédétection sont employées pour produire une classification de l'occupation du sol. Ils sont liés à l'unité de base considérée dans l'analyse des données. Ainsi, nous distinguons les approches basées sur les pixels et celles basées sur les objets (Blaschke, 2010). Dans l'analyse basée sur les pixels, les unités de base sont les pixels de l'image tandis que dans l'approche orientée objet ou OBIA (Object Based Image Analysis), les unités analysées sont les objets. Les objets représentent des regroupements de pixels radiométriquement homogènes et sont extraits préalablement à la classification au moyen d'un algorithme de segmentation. Comparé à un pixel pris individuellement, un objet englobe en plus de ses caractéristiques spectrales plus ou moins homogènes, une dimension ou un contexte spatial (ex. voisinage). L'approche OBIA peut également faciliter le passage à l'échelle du processus de classification en réduisant significativement le nombre d'échantillons d'analyse. Dans cette thèse, nous nous sommes intéressés aux deux paradigmes en proposant une méthode employant chacune des deux approches. Notons également que dans cette thèse, les vérités terrain disponibles sur l'ensemble de nos sites d'étude sont étiquetées de manière spatialement éparse ce qui compromet fortement le recours aux techniques de segmentation sémantique, d'instances ou panoptique en apprentissage profond couramment utilisées ces dernières années pour la cartographie de l'occupation du sol (Audebert et al., 2017; Garnot and Landrieu, 2021; Sirko et al., 2021).

En cartographie de l'occupation du sol, des connaissances spécifiques sur

les classes d'occupation du sol peuvent être mises à profit. Les connaissances à priori susnommées reposent sur une organisation hiérarchisée des classes d'occupation du sol en considérant des relations de classes à sous-classes. À titre d'illustration, un couvert agricole peut être affiné en types de cultures puis les familles de cultures en cultures spécifiques. Le système de classification de l'usage des sols (LCCS) de la FAO (Di Gregorio, 2005) est un exemple concret de taxonomie qui peut être dérivée à partir de classes d'occupation du sol. Ainsi, la prise en compte des relations hiérarchiques caractérisant les classes d'occupation du sol peut s'avérer pertinente dans le processus de classification, notamment dans le cas de l'analyse orientée objet où le nombre d'échantillons d'entraînement est plus réduit par rapport à l'analyse pixel.

Deux méthodes sont proposées dans cette thèse pour la cartographie de l'occupation du sol. La première méthode analyse les séries temporelles radar et optique au niveau objet tout en incorporant dans le processus des connaissances spécifiques sur les classes d'occupation du sol c'est-à-dire leurs relations hiérarchiques. Cette première méthode est basée sur les RNNs qui, rappelons le, sont des approches spécialisées pour la modélisation de données séquentielles. Plus précisément, il est question dans cette méthode d'une extension de la cellule GRU, enrichie avec un mécanisme d'attention modifié afin de mieux tenir compte des spécificités des séries temporelles d'images satellitaires. Une stratégie de pré-entraînement est également proposée pour accompagner le modèle dans la prise en compte des relations hiérarchiques existantes entre classes d'occupation du sol. Enfin, dans l'optique de fournir une contribution sur l'interprétabilité du modèle proposé, nous analysons les poids d'attention et discutons de liens éventuels entre les décisions prises par le modèle et des connaissances notamment agronomiques que nous détenons.

La seconde méthode, quant à elle, explore d'avantage la combinaison de données multi-modales ou multi-sources de télédétection en associant aux séries temporelles radar et optique, une image optique à très haute résolution spatiale. Pour cette seconde méthode, nous considérons plutôt l'utilisation de modèles CNNs qui sont également des approches compétitives pour la classification de données telles que les séries temporelles d'images satellitaires (Pelletier et al., 2019) et présentent l'avantage d'être moins gourmands en temps de calcul que les RNNs pendant la phase d'entraînement. De plus, l'analyse est faite cette fois-ci au niveau pixel sans tenir compte de connaissances à priori sur les classes d'occupation du sol. Il est plus attrayant d'adopter dans ce cas une approche pixel plutôt qu'objet, afin d'explorer différentes architectures de réseaux CNNs (1D, 2D ou 3D), toutes non compatibles avec le niveau objet, en l'occurrence les 2D et 3D. Ainsi, la seconde méthode est basée sur une architecture de CNNs pour laquelle une étude spécifique par source est au préalable menée afin de déterminer le type de modèle CNN le plus conve-

nable. Par ailleurs, la méthode est dotée d'une stratégie d'auto-distillation lui permettant de mieux combiner les sources en entrée.

Ci-après, nous décrirons successivement dans les sections 3.2 et 3.3, chacune des méthodes proposées et y reporterons les résultats de leur évaluation respective.

3.2 Approche HOb2sRNN

Cette section est consacrée à la méthode *HOb2sRNN* (Hierarchical Object based two-Stream Recurrent Neural Network) qui traite de la cartographie de l'occupation du sol à partir de données de télédétection multi-sources publiquement accessibles, en l'occurrence des séries temporelles radar (Sentinel-1) et optique (Sentinel-2), au niveau objet et en incorporant des connaissances à priori sur l'organisation des classes d'occupation du sol. Nous détaillons tout d'abord le fonctionnement de la méthode proposée et présentons ensuite le protocole expérimental ainsi que les résultats de son évaluation.

3.2.1 Description de la méthode

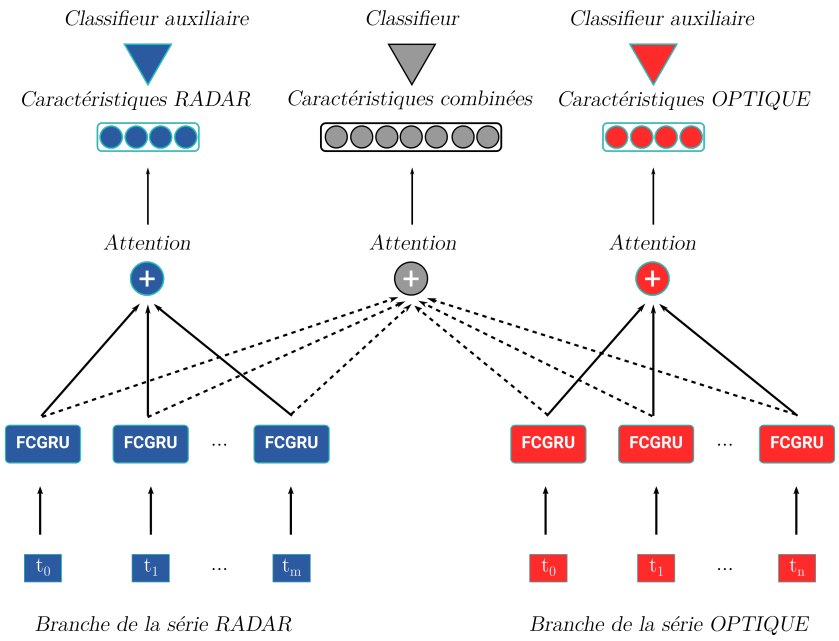


Figure 3.1 – Architecture de la méthode *HOb2sRNN*. Elle a 2 branches, une dédiée à chaque source de données. Chaque branche traite les séries temporelles d'une source au moyen d'une cellule RNN enrichie dénommée FCGRU et d'un mécanisme d'attention employé sur les sorties intermédiaires de la cellule afin d'extraire des caractéristiques par source. De plus, un autre mécanisme d'attention est employé sur la concaténation des sorties intermédiaires par source afin d'obtenir des caractéristiques combinées. Finalement, un classifieur associé aux caractéristiques combinées ainsi que des classifieurs auxiliaires liés aux caractéristiques par branche sont employées pour la prédiction finale.

La Figure 3.1 illustre l'architecture du modèle proposé pour la classification de l'occupation du sol à partir de séries temporelles radar et optique. Elle comprend 2 branches, une pour chaque source. Chaque branche peut être décomposée en 2 parties spécifiques : (i) celle de l'analyse des séries temporelles à travers une extension de la cellule GRU dénommée FCGRU et (ii) celle de la combinaison multi-temporelle des sorties intermédiaires de la cellule FCGRU par le biais d'un mécanisme d'attention modifié afin d'extraire des caractéristiques par source. De plus, un mécanisme d'attention est également employé sur la concaténation des sorties intermédiaires par source afin d'extraire des caractéristiques combinées ou fusionnées. Enfin, l'ensemble des caractéristiques extraites (par source et combinées) est utilisé pour obtenir la classification finale de l'occupation du sol. Par ailleurs, l'apprentissage du modèle est réalisé de sorte à exploiter des connaissances spécifiques au domaine, représentées sous la forme d'une hiérarchie ou taxonomie entre classes d'occupation du sol avec des relations de classes à sous-classes. À cet effet, une stratégie de pré-entraînement du modèle est mise en place, considérant des tâches de complexité graduelle. Ci-après, nous détaillons successivement la cellule FCGRU permettant l'analyse des séries temporelles, le mécanisme d'attention modifié mis en oeuvre dans l'extraction des caractéristiques (par source et combinées), l'emploi de ces caractéristiques pour la classification finale et la stratégie de pré-entraînement du modèle incorporant les relations hiérarchiques entre classes d'occupation du sol.

Cellule FCGRU

La première partie de chaque branche est constituée par une extension directe de la cellule standard GRU présentée en chapitre 1, que nous avons dénommé FCGRU (Fully Connected GRU). Comme illustré sur la figure 3.2, la cellule FCGRU proposée étend la cellule standard GRU en incorporant au début du processus deux couches entièrement connectées (FC_1 et FC_2) avant la transformation héritée de la cellule GRU. Ces couches sont intégrées dans le but de prétraiter les séries temporelles. Elles permettent ainsi d'extraire une combinaison utile des informations d'entrée et d'enrichir la représentation originelle des séries temporelles pour la tâche de classification. Les deux couches s'enchaînent : la couche FC_1 prend en entrée la série temporelle (radar ou optique) et sa sortie est utilisée pour alimenter la couche FC_2. C'est cette représentation enrichie issue de la couche FC_2 qui est utilisée pour la transformation standard héritée de la cellule GRU. Les couches FC_1 et FC_2 sont toutes deux activées par une fonction $tanh$ par souci de cohérence avec le reste des opérations de la cellule basées sur les fonctions $tanh$ et σ. La technique de dropout est employée sur les deux couches entièrement connectées

ainsi que sur la sortie de la cellule afin de prévenir le sur-apprentissage.

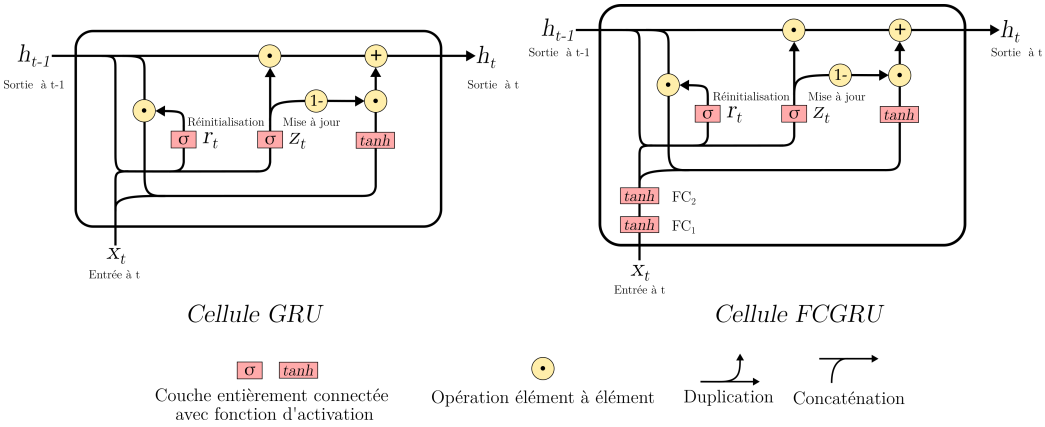


Figure 3.2 – Représentation des cellules GRU et FCGRU

Mécanisme d'attention modifié

La seconde partie des branches consiste en un mécanisme d'attention modifié, combinant les sorties intermédiaires de la cellule FCGRU dans le but d'extraire des caractéristiques par source. Les mécanismes d'attention (voir chapitre 1) sont très utilisés en traitement de signal 1D pour combiner les sorties des RNN au moyen de poids d'attention traditionnellement calculés grâce à une fonction $SoftMax$. Ainsi, les poids d'attention ont des valeurs comprises dans l'intervalle [0,1] et leur somme est contrainte à 1, ce qui apporte une interprétation probabiliste à leur distribution. En raison de la contrainte de somme sur les poids d'attention, il a été remarqué que l'attention $SoftMax$ a la propriété de prioriser une instance (en lui affectant une valeur proche de 1 ce qui écrase le reste des poids) sur les autres, ce qui la rend particulièrement adaptée pour les tâches de traduction automatique où chaque mot cible est généralement aligné sur un des mots sources (Karamanolakis et al., 2019). Cependant, dans le contexte de la classification de l'occupation du sol à partir de séries temporelles d'images satellitaires, il est rare qu'une seule estampille temporelle soit déterminante aux dépens de toutes les autres. Par conséquent, forcer la somme des poids à 1 peut ne pas être approprié dans notre contexte où plusieurs estampilles temporelles peuvent se révéler pertinentes pour les diverses classes d'occupation du sol. C'est pourquoi dans la formulation du mécanisme d'attention adopté pour le modèle proposé, nous avons relaxé cette contrainte en substituant la fonction $SoftMax$ par une fonction $tanh$ dans le calcul des poids d'attention. Ceci permet de pondérer les estampilles temporelles indépendamment les unes

des autres tel dans un système à portes comme celui adopté pour contrôler le flux d'information dans les RNNs avancés (LSTM, GRU). La fonction $tanh$ autorise par ailleurs une gamme de valeurs plus large pour les poids d'attention avec des occurrences négatives, à savoir [-1,1]. Les équations suivantes définissent formellement le mécanisme d'attention introduit, modifié avec la fonction $tanh$:

$$score = v\ tanh(Wh_t + b) \tag{3.1}$$

$$\alpha = tanh(score) \tag{3.2}$$

$$feat = \sum_i \alpha_i h_{t_i} \tag{3.3}$$

où h_t représente les sorties intermédiaires de la cellule FCGRU tandis que la matrice W et les vecteurs v et b sont des paramètres appris par le modèle. α est le poids d'attention traditionnellement calculé à l'aide d'une fonction $SoftMax$ remplacée ici par une fonction $tanh$. Notons que dans le cas présent de la classification des séries temporelles d'images satellitaires où il n'y a nul besoin de décodeur comme c'est le cas pour l'architecture encodeur-décodeur, le vecteur de contexte sert directement de caractéristiques ($feat$) pour la classification.

Le mécanisme d'attention ci-dessus décrit est employé pour combiner les sorties intermédiaires du FCGRU dans chacune des branches et extraire des caractéristiques par source encodant l'information temporelle en entrée. Ces dernières sont dénommées $feat_{rad}$ et $feat_{opt}$ respectivement pour la branche radar et optique. En plus, un autre mécanisme d'attention est également employé sur la concaténation des sorties intermédiaires par source du FC-GRU afin d'extraire des caractéristiques fusionnées. Celles-ci sont dénommées $feat_{fused}$ et encodent à la fois l'information des séries temporelles radar et optique mais également leur complémentarité. Nous nommons également α_{rad}, α_{opt} et α_{fused}, l'ensemble des poids d'attention associés respectivement à la combinaison des sorties intermédiaires de la branche radar, de la branche optique et de leur concaténation.

Emploi des caractéristiques pour la classification finale

L'ensemble des caractéristiques, une fois généré, est employé pour effectuer la classification finale de l'occupation du sol. Trois classifieurs sont

utilisés à cet effet : un classifieur associé aux caractéristiques issues de la combinaison des sorties intermédiaires des branches radar et optique et deux classifieurs auxiliaires associées aux caractérisitiques par source issues des branches. Le classifieur correspondant aux caractéristiques fusionnées est un réseau de neurones constitué de deux couches entièrement connectées avec une activation ReLU et la technique de dropout, suivies d'une couche de sortie avec une activation $SoftMax$. Quant aux classifieurs auxiliaires, elles correspondent chacune à une couche de sortie avec une activation $SoftMax$. Les classifieurs auxiliaires sont utilisés pour renforcer la complémentarité entre les caractéristiques extraites par source et les rendre chacune, le plus possible, discriminantes par elles-mêmes indépendamment des autres (Hou et al., 2017; Interdonato et al., 2019; Ienco et al., 2019b). La perte (L) servant à l'optimisation des trois classifieurs est définie comme suit :

$$L = CE(Y, CL(feat_{fused})) + \lambda \sum_{source \in \{rad,opt\}} CE(Y, OUT(feat_{source})) \tag{3.4}$$

où CE est l'entropie croisée dans un contexte multi-classe définie par l'équation (3.6) ; Y est l'information de référence c'est-à-dire la classe d'occupation du sol à prédire ; CL correspond au classifieur associé aux caractéristiques fusionnées et OUT est un classifieur auxiliaire. Enfin, λ est un hyper-paramètre qui pondère le coût lié aux classifieurs auxiliaires. La prédiction finale de l'occupation du sol ($pred$) est obtenue en combinant les sorties des trois classifieurs suivant le schéma de pondération adopté précédemment :

$$pred = arg\ max(CL(feat_{fused}) + \lambda \sum_{source \in \{rad,opt\}} OUT(feat_{source})) \tag{3.5}$$

$$CE = -\sum_{c=1}^{C} y_c\ log(p_c) \tag{3.6}$$

où C est le nombre de classes ; y_c correspond à l'étiquette à prédire et p_c est la distribution de probabilité $SoftMax$ en sortie des classifieurs.

Stratégie de pré-entraînement

Dans le but d'incorporer les connaissances spécifiques sur les classes d'occupation du sol dans l'apprentissage du modèle *HOb2sRNN*, nous avons mis en place une stratégie de pré-entraînement des paramètres qui y sont associés.

Celle-ci consiste à répéter l'apprentissage du modèle pour chaque niveau de l'organisation hiérarchique des classes d'occupation du sol, du plus général au plus spécifique ou au plus fin qui correspond également au niveau de classification envisagé. Plus précisément, l'apprentissage du modèle est initialisé au niveau le plus élevé de la taxonomie, puis se poursuit au niveau suivant en réutilisant les paramètres appris précédemment pour l'ensemble de l'architecture, excepté ceux des classifieurs, puisque ces derniers sont spécifiques à chaque niveau (Figure 3.3). Ainsi, de nouveaux paramètres sont appris à chaque niveau de la hiérarchie pour l'ensemble des classifieurs. Cette stratégie est conduite jusqu'à ce que le niveau de classification cible soit atteint. En résumé, la stratégie de pré-entraînement hiérarchisée introduite permet au modèle de se focaliser en premier sur des problèmes de classification avec un haut niveau d'abstraction et progressivement, de s'adapter à des tâches de classification de complexité croissante. De plus, le processus mis en place permet au modèle d'aborder la classification au niveau cible en intégrant une sorte de connaissance préalable de la tâche (basée sur les classes de haut niveau) au lieu de l'aborder de bout-en-bout à partir de rien.

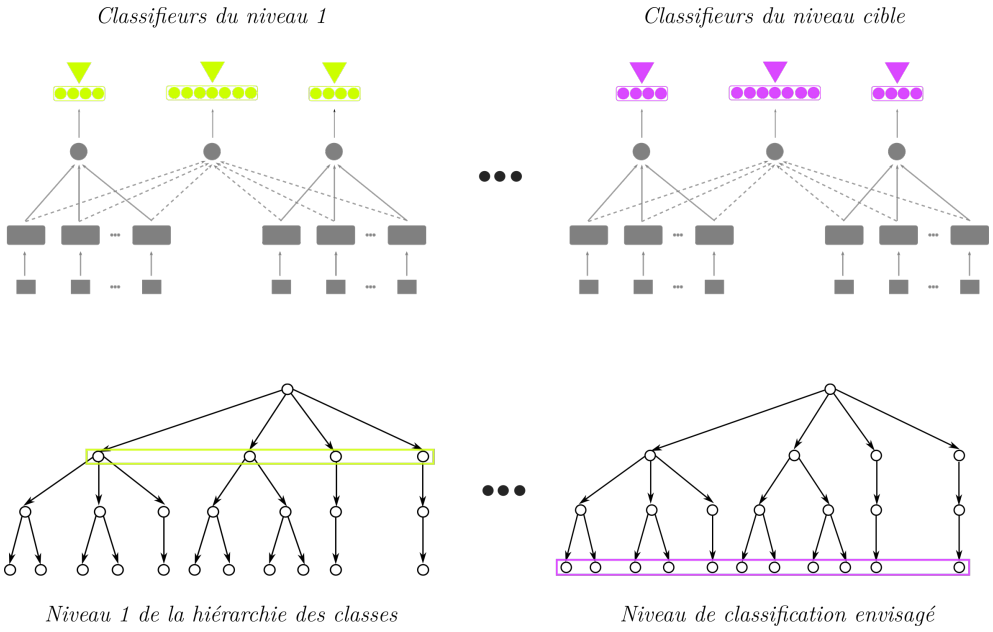


Figure 3.3 – Illustration de la stratégie de pré-entraînement du modèle

3.2.2 Protocole expérimental

L'évaluation de la méthode *HOb2sRNN* est conduite sur les sites de l'île de la Réunion et du bassin arachidier au Sénégal. Les dates d'acquisition des séries temporelles Sentinel-1 et Sentinel-2 sont illustrées par les figures 3.4 et 3.5, respectivement pour l'île de la Réunion et le site du Sénégal. Notons que pour les données Sentinel-2, seules les bandes spectrales à 10-m sont considérées en plus de l'indice de végétation NDVI.

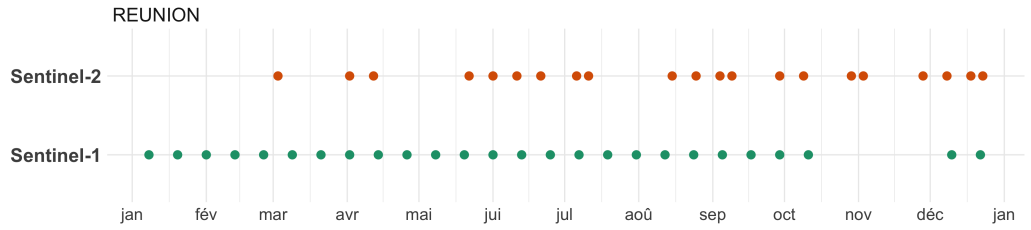


Figure 3.4 – Dates d'acquisition des 26 images Sentinel-1 et 21 images Sentinel-2 sur l'île de la Reunion

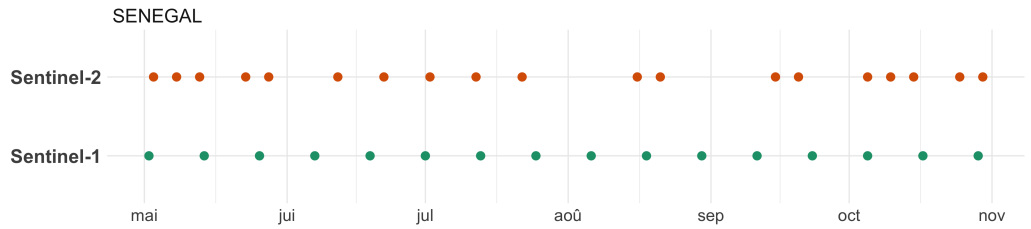


Figure 3.5 – Dates d'acquisition des 16 images Sentinel-1 et 19 images Sentinel-2 sur le site du Sénégal

Dans le but de conduire l'analyse des données au niveau objet, une segmentation est effectuée au préalable sur chaque site d'étude. Pour ce faire, nous utilisons les images à très haute résolution spatiale disponibles sur chacun des sites d'étude c'est-à-dire une image SPOT-6/7 sur l'île de la Réunion et une image PlanetScope sur le site du Sénégal. Afin d'assurer une correspondance spatiale précise, chacune des images est au préalable recalée sur la grille des données Sentinel à partir de points homologues extraits automatiquement. Elles sont ensuite segmentées avec l'algorithme Large Scale Generic Region Merging (LSGRM) implémenté dans l'Orfeo Toolbox (Lassalle et al., 2015). Les paramètres de segmentation (seuils d'homogénéité spatiale et spectrale) sont ajustés par interprétation visuelle à travers plusieurs essais afin que les objets finaux segmentés correspondent le plus possible aux unités

d'occupation du sol des sites étudiés. Un aperçu des couches de segmentation obtenues sur les deux sites est illustré par les figures 3.6 et 3.7.

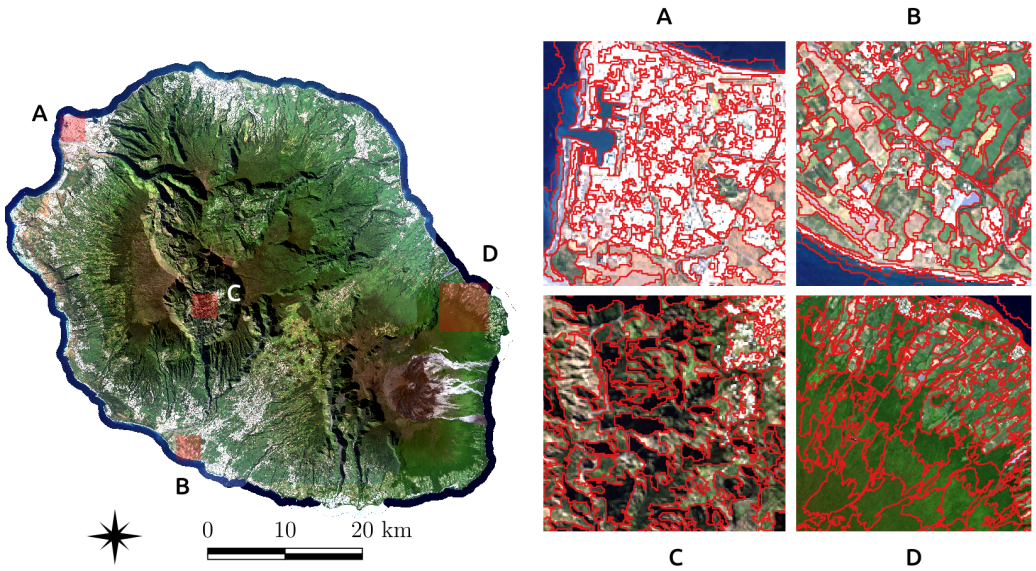


Figure 3.6 – Quelques aperçus de la couche de segmentation obtenue sur l'île de la Réunion

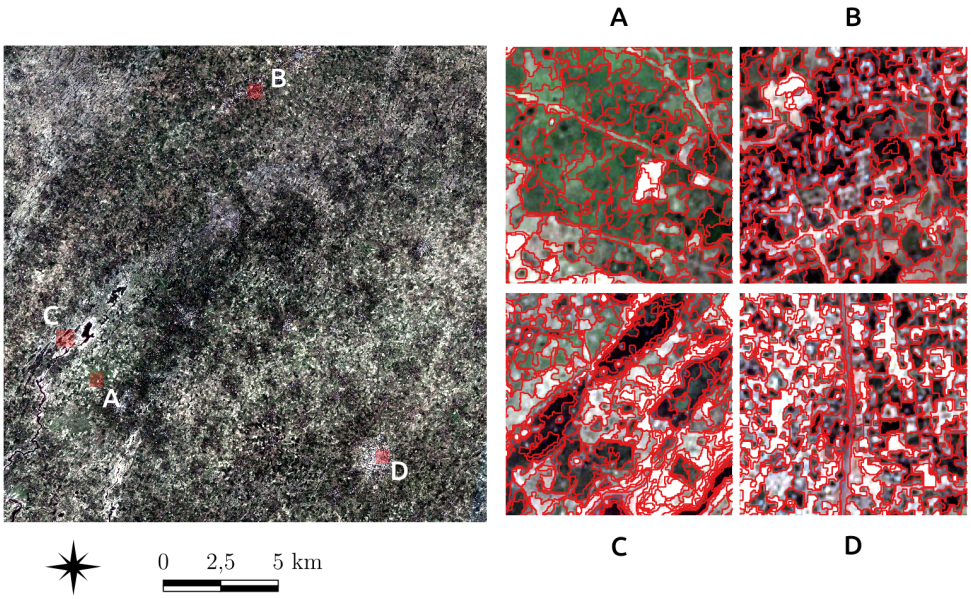


Figure 3.7 – Quelques aperçus de la couche de segmentation obtenue sur le site du Sénégal

Les vérités terrain disponibles en format vecteur sur les deux sites contiennent une collection de polygones, chacune appartenant à une classe d'occupation du sol spécifique. Pour rappel, 6 265 et 734 polygones labélisés ont été respectivement collectés sur l'île de la Réunion et le site du Sénégal (voir chapitre 2). Afin d'obtenir une base d'apprentissage constituée d'échantillons radiométriquement homogènes sur les deux sites d'études, les vérités terrain ont été spatialement intersectées avec les couches de segmentation produites. Ainsi 7 908 et 3 084 nouveaux polygones labélisés sont finalement obtenus, respectivement sur l'île de la Réunion et le site du Sénégal (Tableaux 3.1 et 3.2). Outre l'obtention d'échantillons radiométriquement homogènes, ce procédé a pour intérêt d'augmenter sensiblement le nombre total d'échantillons de la base d'apprentissage, en l'occurrence sur le site du Sénégal (un peu plus de 4 fois). Enfin, les valeurs moyennes des objets sont extraites des séries temporelles pour permettre l'apprentissage des modèles au niveau orienté-objet.

Tableau 3.1 – Nombre de polygones par classe sur l'île de la Réunion

Classe	Polygones
1 – CANNE À SUCRE	1 258
2 – PÂTURAGE ET FOURRAGE	869
3 – MARAÎCHAGE	912
4 – CULTURE SOUS SERRE OU OMBRAGÉE	233
5 – ARBORICULTURE	1 014
6 – ESPACE BOISE	1 106
7 – LANDE ET SAVANE	850
8 – ROCHER ET SOL NU NATUREL	573
9 – OMBRE DUE AU RELIEF	107
10 – EAU	261
11 – ESPACE ARTIFICIALISÉ	725
TOTAL	7 908

Afin d'injecter des connaissances à priori sur les classes d'occupation du sol dans l'apprentissage, nous avons dérivé sur les deux sites d'étude, une organisation hiérarchique des classes d'occupation du sol. (Figures 3.8 et 3.9). Les taxonomies mises en place comprennent trois niveaux de représentation dont le plus fin correspond au niveau de classification cible décrit dans les tableaux 3.1 et 3.2.

Avec le dessein de fournir une analyse approfondie du modèle *HOb2sRNN*, nous avons évalué ses performances par rapport à cinq autres approches d'état de l'art dans la cartographie de l'occupation du sol. La première approche adoptée est celle des forêts aléatoires (RF) évaluée selon deux stratégies de fusion des données : (i) une fusion précoce où les informations radar et

Tableau 3.2 – Nombre de polygones par classe sur le site du Sénégal

Classe	Polygones
1 – FORMATION BUISSONNANTE	100
2 – JACHÈRE ET FRICHE	322
3 – MARE	59
4 – RIVE ET SOL NU	132
5 – VILLAGE	767
6 – ZONE HUMIDE	156
7 – VALLÉE	56
8 – CÉRÉALE	816
9 – LÉGUMINEUSE	676
TOTAL	3 084

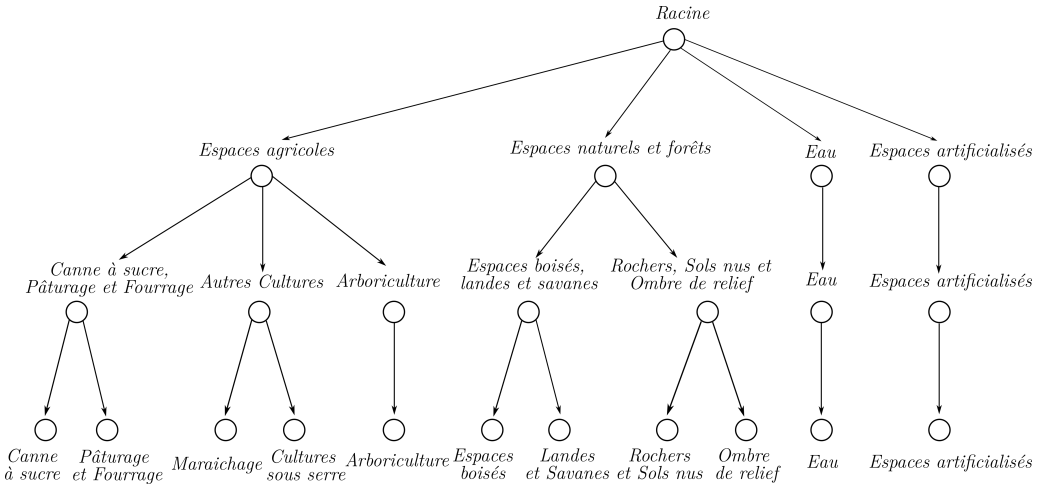


Figure 3.8 – Hiérarchie des classes d'occupation du sol sur l'île de la Réunion

optique sont concaténées pour entraîner le classifieur (Erinjery et al., 2018) ; par la suite cette variante est désignée RF_{early} et (ii) une fusion tardive dans laquelle un classifieur est entraîné par source, puis la décision finale obtenue par le produit des sorties respectives par source (Valero et al., 2019) ; cette variante est dénommée RF_{late}. Les autres approches sont : un classifieur SVM (Support Vector Machine) ; un perceptron multi-couche (MLP) comme celui employé comme classifieur des caractéristiques fusionnées du modèle *HOb2sRNN* ; le modèle TempCNN (Temporal Convolutional Neural Network) proposé par Pelletier et al. (2019) qui procède par des convolutions temporelles 1D et le modèle OD2RNN (Ienco et al., 2019a) issu des recherches préliminaires qui ont abouti à la méthode *HOb2sRNN*. Tout comme pour les approches RF, les modèles SVM et MLP sont entraînés en employant la

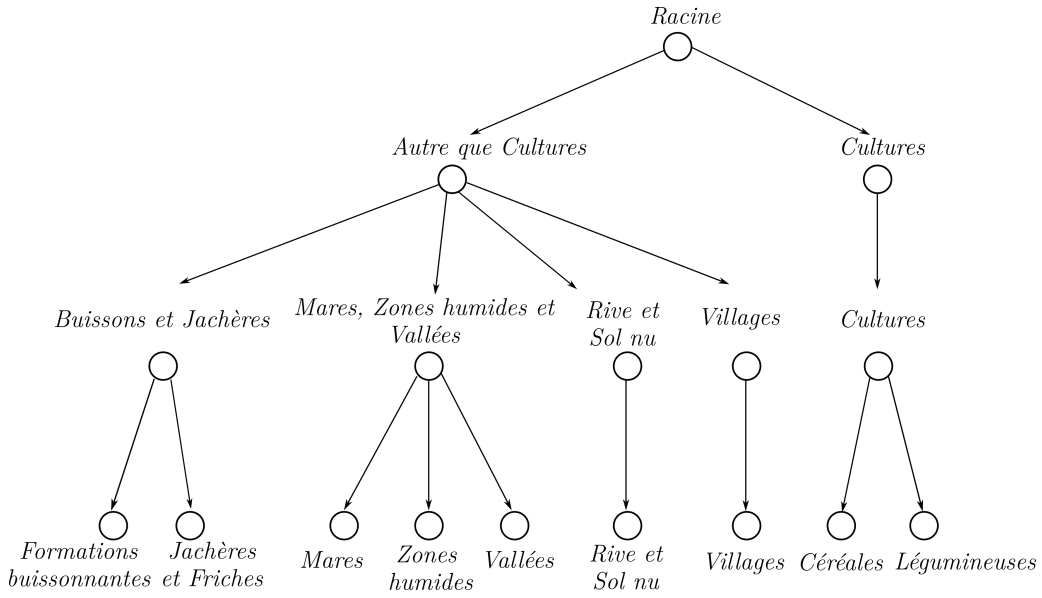


Figure 3.9 – Hiérarchie des classes d'occupation du sol sur le site du Sénégal

concaténation des informations radar et optique. Pour l'approche TempCNN, nous avons mis en place un modèle à 2 branches, une par source de données. L'architecture de chaque branche reste identique à celle décrite par Pelletier et al. (2019). Enfin, le modèle OD2RNN est employé avec la même structure et les mêmes paramètres pris en compte par Ienco et al. (2019a). Le nombre de paramètres optimisables des approches par réseaux de neurones est reporté dans le tableau 3.3.

De plus, nous avons mené une analyse de sensibilité sur l'hyper-paramètre λ du modèle *HOb2sRNN* pondérant la contribution des classifieurs auxiliaires et dont la valeur est initialement fixée à 0.5. Une étude d'ablation est également conduite sur les sources d'entrée ainsi que sur les principales composantes du modèle afin de démêler leurs avantages au regard de la méthode proposée. Trois composantes sont exclues tour à tour. En considérant les données multi-sources, nous faisons tout d'abord abstraction des mécanismes d'attention utilisés. Cette variante du modèle est dénommée $HOb2sRNN_{NoAtt}$). Ensuite, nous faisons fi de la stratégie de pré-entraînement hiérarchisée adoptée. Cette dernière est dénommée $HOb2sRNN_{NoPreHier}$. Pour finir, nous négligeons l'étape d'enrichissement liée à l'emploi des deux couches entièrement connectées dans la cellule FCGRU. Ceci équivaut à l'utilisation d'une cellule GRU. Cette variante est nommée $HOb2sRNN_{GRU}$. Nous avons par ailleurs examiné une autre variante du modèle, où nous revenons au mécanisme d'attention classique employant une fonction $SoftMax$. Elle est dénommée

$HOb2sRNN_{SoftMaxAtt}$. Enfin, deux analyses qualitatives sont fournies : la première porte sur les cartes d'occupation du sol produites par les différentes méthodes tandis que la seconde consiste en une inspection des poids d'attention α appris par le modèle *HOb2sRNN*, en vue d'examiner dans quelle mesure ils peuvent contribuer à son interprétabilité.

Tableau 3.3 – Nombre de paramètres optimisables des approches par réseaux de neurones sur les deux sites

Modèles	Paramètres optimisables	
	Réunion	Sénégal
MLP	349 195	332 809
TempCNN	465 739	268 617
OD2RNN	2 173 761	2 160 667
HOb2sRNN	4 391 810	4 382 576

Les valeurs des données d'apprentissage sont normalisées par bande, considérant les séries temporelles mulit-source, dans l'intervalle [0,1]. Les données sont divisées en jeu d'entraînement, jeu de validation et jeu test avec des proportions respectives de 50%, 20% et 30% des objets labélisés. Pour ce faire, nous nous assurons également que tous les objets labélisés, issus du même polygone de vérité terrain avant l'intersection spatiale, se retrouvent exclusivement dans l'une des 3 partitions (entraînement, validation ou test) afin d'éviter au mieux de possibles biais d'auto-corrélation spatiale dans la procédure d'évaluation. Par ailleurs, les modèles sont optimisés à travers une procédure d'entraînement et validation : les hyper-paramètres des modèles RF et SVM sont variés dans une gamme de valeurs afin de sélectionner les valeurs leur permettant de généraliser au mieux pendant l'entraînement sur le jeu de validation ; dans le cas des réseaux de neurones (MLP, TempCNN, OD2RNN et *HOb2sRNN*), les valeurs des hyper-paramètres sont fixées dans l'élaboration des architectures et ce sont les paramètres ou poids permettant d'obtenir une meilleure généralisation sur le jeu de validation pendant les époques d'entraînement qui sont sauvegardés. Les hyper-paramètres et valeurs associées des méthodes sont reportés dans le tableau 3.4.

Trois métriques sont considérées pour évaluer les performances des modèles sur le jeu test : la précision globale, le F1-score et le cœfficient Kappa (voir Section 1.2.3 pour les définitions). Étant donné que les performances des modèles que nous évaluons, sont susceptibles de varier en fonction de la division des données, en raison d'échantillons plus simples ou plus complexes impliqués dans les différentes partitions, ces métriques sont moyennées sur 10 divisions aléatoires du jeu de données suivant la stratégie précédemment décrite. Les expériences sont conduites sur une station de travail dotée d'un

CPU Intel Xeon, d'une RAM de 256 GB et de 4 cartes GPU NVIDIA TITAN X. Avec une telle configuration, l'entraînement du modèle *HOb2sRNN* a duré approximativement 16 et 4.5 heures, respectivement sur l'île de la Réunion et le site du Sénégal. L'implémenation du modèle *HOb2sRNN* est disponible à l'adresse https://github.com/eudesyawog/HOb2sRNN.

Tableau 3.4 – Hyper-paramètres et valeurs associées des compétiteurs

Méthode	**Hyper-paramètre**	**Valeur ou Gamme**
RF	Nombre d'arbres	{100, 200, 300,400,500}
	Profondeur maximale	{20,40,60,80,100}
	Maximum de caractéristiques	{'sqrt', 'log2', aucune}
SVM	Noyau	{'linéaire', 'polynomial', 'RBF', sigmoïde'}
	Cœfficient γ	{0.25,0.5,1,2}
	Pénalité	{0.1, 1, 10}
MLP	Unités par couche	512
	Couches cachées	2
	Taux de Dropout	0.4
HOb2sRNN	Unités FCGRU	512
	Unités FC_1	64
	Unités FC_2	128
	Unités du classifieur des caractéristiques combinées	512 par couche
	Taux de Dropout	0.4
	Hyper-paramètre λ	0.5
Tous les réseaux de neurones	Taille par lot	32
	Optimiseur	Adam (Kingma and Ba, 2015)
	Taux d'apprentissage	10^{-4}
	Nombre d'époques	2000 (par niveau pour *HOb2sRNN*)

3.2.3 Évaluation quantitative

Performances générales des modèles

Nous reportons dans le tableau 3.5, les performances moyennes des modèles sur les deux sites d'étude. Nous remarquons que la méthode proposée performe mieux que ses compétiteurs sur les deux sites d'étude, bien que l'écart de performance soit plus prononcé sur le jeu de données de la Réunion que sur le site du Sénégal. Ceci peut être dû au fait que le jeu de données de l'île de la Réunion possède plus d'échantillons de vérité terrain (environ

Tableau 3.5 – Performances moyennes des modèles sur les deux sites d'étude

Reunion	F1-score	Kappa	Précision globale
RF_{early}	75.62 ± 1.00	0.726 ± 0.011	75.75 ± 0.98
RF_{late}	74.26 ± 0.75	0.713 ± 0.009	74.72 ± 0.78
SVM	75.34 ± 0.88	0.722 ± 0.010	75.39 ± 0.89
MLP	77.96 ± 0.70	0.752 ± 0.008	78.03 ± 0.66
TempCNN	77.76 ± 1.06	0.749 ± 0.012	77.79 ± 1.05
OD2RNN	74.39 ± 1.14	0.712 ± 0.012	74.50 ± 1.09
HOb2sRNN	**79.66** ± 0.85	**0.772** ± 0.009	**79.78** ± 0.82
Sénégal	F1-score	Kappa	Précision globale
RF_{early}	86.31 ± 0.91	0.828 ± 0.012	86.35 ± 0.90
RF_{late}	85.31 ± 0.50	0.816 ± 0.006	85.45 ± 0.48
SVM	89.95 ± 0.85	0.875 ± 0.011	89.96 ± 0.85
MLP	90.05 ± 0.56	0.876 ± 0.007	90.07 ± 0.57
TempCNN	88.81 ± 0.58	0.861 ± 0.007	88.83 ± 0.58
OD2RNN	88.35 ± 0.72	0.855 ± 0.009	88.34 ± 0.72
HOb2sRNN	**90.78** ± 1.03	**0.885** ± 0.013	**90.78** ± 1.03

huit fois) que le jeu de données sénégalais. Il est généralement admis que les modèles d'apprentissage profond parviennent à être plus performants que les méthode classiques d'apprentissage automatique, lorsqu'ils sont entraînés sur d'importants volumes de données. En ce qui concerne les autres approches évaluées, en particulier les modèles RF (RF_{early} et RF_{late}), nous notons que la stratégie de fusion tardive c'est-à-dire RF_{late} s'avère moins efficace que la concaténation directe des deux sources de données. L'approche SVM obtient des scores similaires à ceux de RF_{early} sur l'île de la Réunion, tandis qu'elle surpasse ce dernier sur le site du Sénégal. Ce dernier point met particulièrement en évidence la pertinence de l'algorithme SVM pour les jeux de données limités en termes d'échantillons étiquetés. Quant aux modèles MLP et TempCNN, tous deux ont obtenu des scores inférieurs à ceux de *HOb2sRNN* sur l'île de la Réunion, tandis que les performances du modèle MLP sont comparables à celles de *HOb2sRNN* sur le site du Sénégal. De plus, les performances du modèle OD2RNN sur les deux sites étudiés indiquent la plus-value des extensions apportées par le modèle *HOb2sRNN*. Pour finir, notons que les scores, relativement meilleurs, obtenus sur le site du Sénégal par rapport à celui de la Réunion proviennent sans doute de l'aspect topographique liée à ces deux sites. En effet, l'île de la Réunion est caractérisée par une topographie accidentée alors que le site sénégalais est essentiellement plat. Les effets de relief, comme l'ombre ou l'orientation, peuvent induire des biais dans la discrimination des classes d'occupation du sol impactant beaucoup plus certainement l'île de la Réunion (Ienco et al., 2019b).

Analyse sur les classes d'occupation du sol

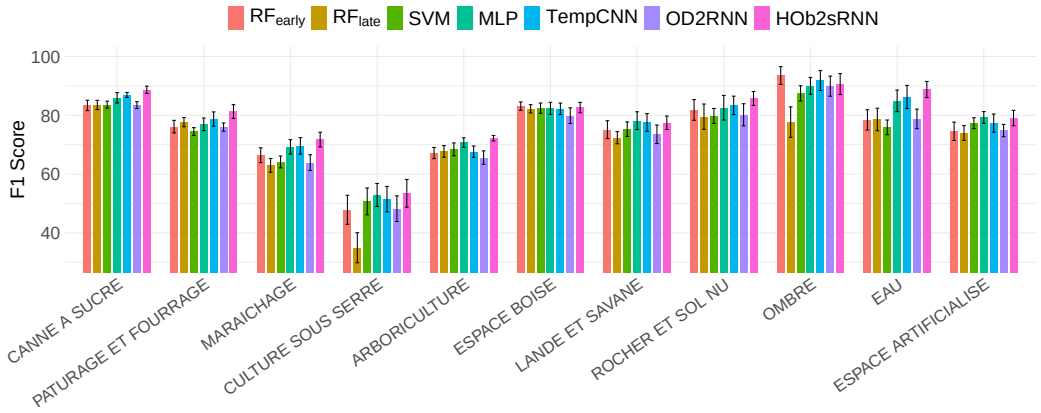


Figure 3.10 – Performances moyennes par classe d'occupation du sol sur l'île de la Réunion (l'écart-type est affichée sous forme de barre d'erreur)

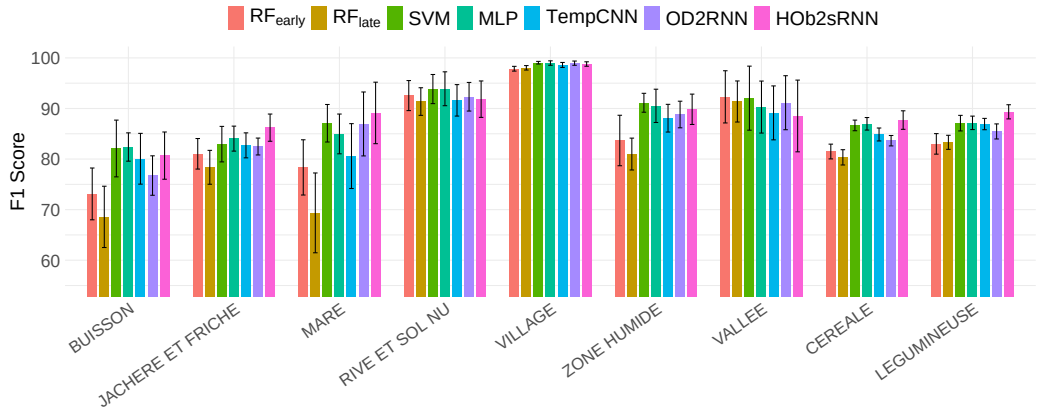


Figure 3.11 – Performances moyennes par classe d'occupation du sol sur le site du Sénégal (l'écart-type est affichée sous forme de barre d'erreur)

La suite de notre évaluation est portée sur l'analyse des performances par classe d'occupation du sol. Les valeurs de F1-score par classe sur les deux sites sont illustrés respectivement par les figures 3.10 et 3.11. Sur le site de la Réunion, nous notons que le modèle proposé obtient de meilleures performances sur une majorité des classes d'occupation du sol, excepté quelques unes où ses performances restent cependant compétitives par rapport aux autres approches comme RF ou MLP. Il convient de noter également que le modèle *HOb2sRNN* est particulièrement efficace sur les classes d'occupation

du sol relatifs à l'agriculture ou à la végétation en l'occurrence la canne à sucre, les pâturage et fourrage, les cultures maraîchères ou aussi l'arboriculture. Ceci souligne le fait que la méthode proposée est bien adaptée pour tirer parti des dépendances temporelles caractérisant ces classes d'occupation du sol. Sur le site du Sénégal, les performances par classe d'occupation du sol sont plus modérées pour le modèle *HOb2sRNN*. Elle a toutefois obtenu de meilleurs scores sur les classes jachère et friche, mare, céréale et légumineuse. Il convient également de noter dans ce cas, que les classes les mieux caractérisées par le modèle proposé, présentent des dynamiques temporelles notables. Par exemple, il est courant d'observer de la végétation naturelle pousser sur les zones en jachère ; les mares se créent le plus souvent pendant la saison pluvieuse ; ou encore les classes céréale et légumineuse qui ont un cycle de développement lié à la saison pluvieuse. Ces résultats sous-tendent les observations faites sur l'île de la Réunion et renforcent le fait que la méthode proposée est capable de tirer parti des dépendances temporelles pour prendre ses décisions.

Pour approfondir l'analyse par classe, nous avons également examiné les erreurs de classification commises par l'ensemble des modèles sur les deux sites d'étude. Les matrices de confusion entre classes sont reportées respectivement sur les figures 3.12 et 3.13. Sur l'île de la Réunion, nous observons particulièrement des erreurs de classification entre espace artificialisé et culture sous serre généralisées pour l'ensemble des méthodes. Pour le reste, les résultats sont cohérents avec les scores par classe précédemment discutés. Sur le site du Sénégal, les confusions varient sensiblement entre méthodes. Par exemple, les approches RF présentent des confusions pour les formations buissonnantes et jachères qui sont respectivement mal classés en céréale et légumineuse. Nous pouvons également souligner la confusion entre mare et zone humide. Les autres méthodes, notamment la méthode proposée, ont tendance à réduire ces confusions comme le soulignent leur matrices respectives.

Analyse de sensibilité sur la pondération liée aux classifieurs auxiliaires

Dans cette partie, nous analysons de quelle façon l'hyper-paramètre λ, associé à la contribution des classifieurs auxiliaires, influence les performances du modèle proposé. À cette fin, nous faisons varier λ, initialement fixé à 0.5, dans l'intervalle [0.1, 0.7] en incrémentant sa valeur de 0.1. Les résultats présentés en figure 3.14 montrent de manière générale une stabilité des performances obtenues sur les deux sites d'étude. De plus, dans la gamme de valeur adoptée pour l'hyper-paramètre λ, les performances du modèle *HOb2sRNN* restent compétitives par rapport à celles des autres approches évaluées précé-

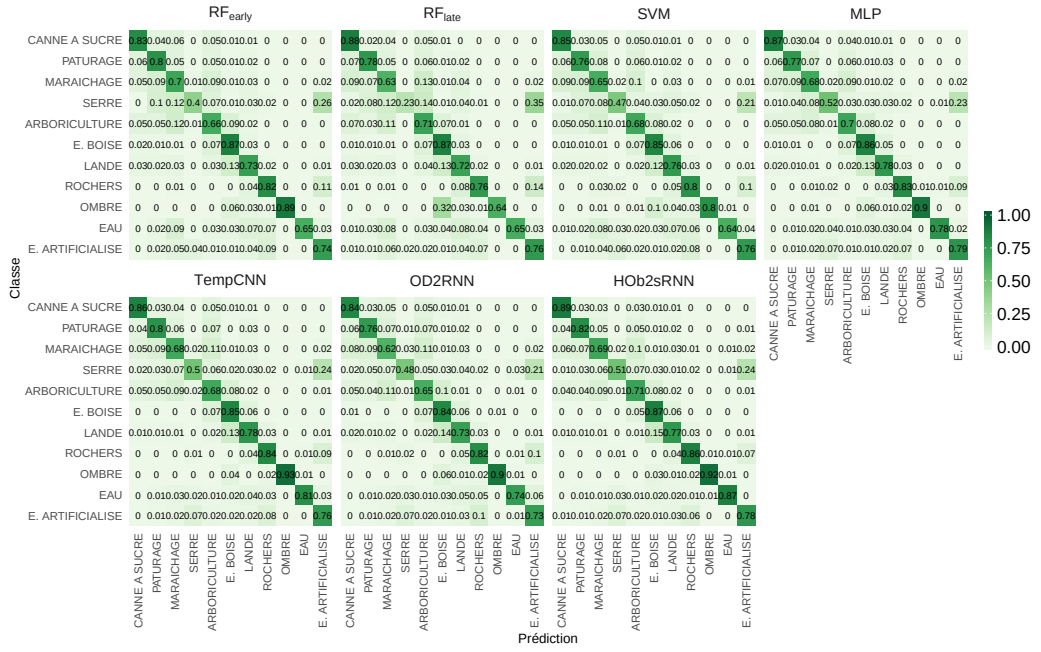


Figure 3.12 – Matrices de confusion entre les classes d'occupation du sol de l'île de la Réunion

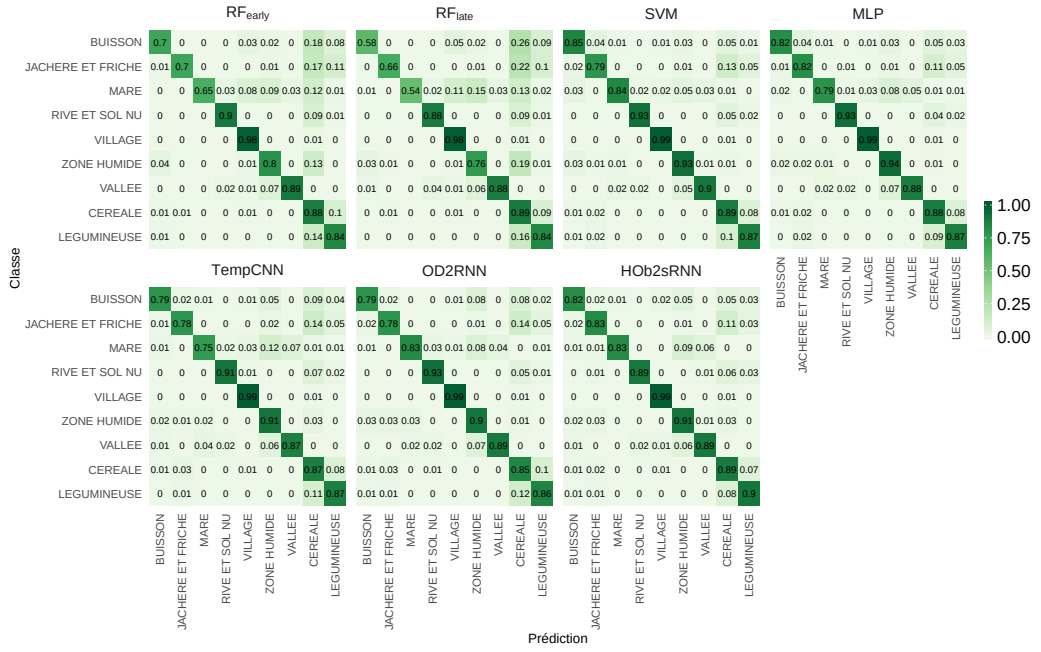


Figure 3.13 – Matrices de confusion entre les classes d'occupation du sol du site du Sénégal

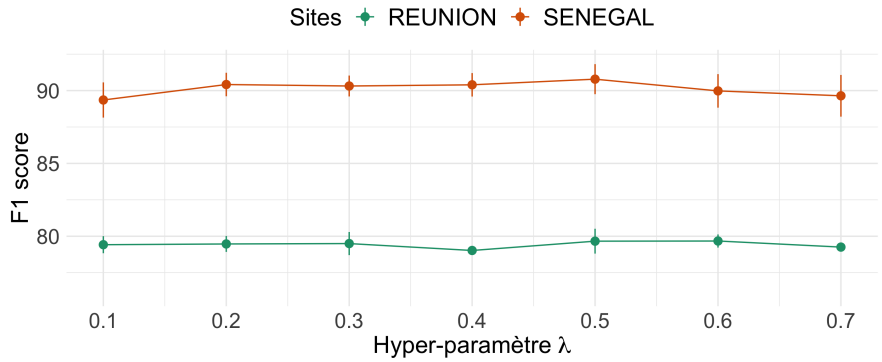


Figure 3.14 – Analyse de sensibilité du modèle *HOb2sRNN* à la variation de l'hyper-paramètre λ associé à la perte des classifieurs auxiliaires. L'écart-type est affiché sous forme de barre d'erreur.

demment. Nous notons que le F1-score, sur le site de la Réunion, varie entre 79.02 et 79.87% pour des valeurs λ respectives de 0.4 et 0.6, tandis que sur le site du Sénégal, celui-ci varie entre 89.35 lorsque λ prend la valeur 0.1 et 90.78% lorsque λ est égal à 0.5.

Analyse d'ablation sur les sources d'entrée

Dans cette étape de notre analyse quantitative, nous réalisons une étude d'ablation sur les données multi-sources en entrée des différents modèles évalués. Une seule source à la fois est alors considérée pour effectuer la classification de l'occupation du sol, entraînant l'utilisation d'une branche unique pour les modèles TempCNN, OD2RNN et *HOb2sRNN*. Les résultats sont reportés respectivement dans les tableaux 3.6 et 3.7 pour les deux sites.

Nous remarquons tout d'abord un comportement spécifique de la source radar (Sentinel-1) à l'égard de chacun des sites. Si la source radar est plus ou moins discriminante sur le site du Sénégal, il n'en va pas de même pour l'île de la Réunion, compte tenu des faibles performances obtenues pour les méthodes évaluées notamment l'approche SVM. Comme mentionné précédemment, l'île de la Réunion est caractérisée par un relief très accidenté comparé au site du Sénégal situé dans une vaste plaine. Le signal radar étant beaucoup plus sensible que le signal optique aux effets de relief élevé, les performances des modèles évalués ont certainement été négativement impactées sur le site de la Réunion lorsque seule la source radar était employée. Ainsi, lorsque les deux sources sont combinées (voir Tableau 3.5), la majorité des approches ont des performances légèrement moins bonnes ou similaires à celles de l'emploi de la source optique seule (Sentinel-2), en raison du bruit

Tableau 3.6 – Performances moyennes des modèles évalués sur le site de la Réunion considérant l'ablation de l'une des sources

Sentinel-1	F1-score	Kappa	Précision globale
RF	**36.77** ± 0.93	**0.291** ± 0.011	**37.85** ± 0.95
SVM	6.56 ± 0.36	0.018 ± 0.009	16.85 ± 0.53
MLP	34.93 ± 1.42	0.271 ± 0.016	36.01 ± 1.39
TempCNN	32.28 ± 1.19	0.239 ± 0.013	33.17 ± 1.17
OD2RNN	31.83 ± 0.98	0.234 ± 0.012	32.71 ± 1.01
HOb2sRNN	31.80 ± 1.10	0.231 ± 0.011	32.39 ± 1.04
Sentinel-2	F1-score	Kappa	Précision globale
RF	76.24 ± 0.59	0.732 ± 0.007	76.32 ± 0.63
SVM	75.55 ± 0.80	0.724 ± 0.009	75.60 ± 0.80
MLP	77.95 ± 0.69	0.751 ± 0.008	77.98 ± 0.73
TempCNN	78.25 ± 0.88	0.755 ± 0.010	78.27 ± 0.90
OD2RNN	74.55 ± 0.81	0.714 ± 0.008	74.66 ± 0.72
HOb2sRNN	**78.69** ± 0.95	**0.761** ± 0.010	**78.79** ± 0.91

Tableau 3.7 – Performances moyennes des modèles évalués sur le site du Sénégal considérant l'ablation de l'une des sources

Sentinel-1	F1-score	Kappa	Précision globale
RF	75.71 ± 1.03	0.703 ± 0.013	76.56 ± 1.00
SVM	71.27 ± 0.82	0.653 ± 0.010	72.82 ± 0.78
MLP	**78.96** ± 1.28	**0.738** ± 0.015	**79.05** ± 1.23
TempCNN	77.79 ± 0.79	0.725 ± 0.010	78.01 ± 0.80
OD2RNN	75.07 ± 1.59	0.692 ± 0.019	75.34 ± 1.50
HOb2sRNN	77.42 ± 1.33	0.721 ± 0.016	77.63 ± 1.27
Sentinel-2	F1-score	Kappa	Précision globale
RF	84.51 ± 1.17	0.806 ± 0.015	84.60 ± 1.17
SVM	**88.64** ± 0.47	**0.858** ± 0.006	**88.63** ± 0.45
MLP	88.38 ± 0.61	0.855 ± 0.008	88.40 ± 0.62
TempCNN	87.42 ± 1.02	0.843 ± 0.013	87.42 ± 1.04
OD2RNN	86.03 ± 0.75	0.826 ± 0.010	86.01 ± 0.75
HOb2sRNN	87.56 ± 1.33	0.845 ± 0.017	87.55 ± 1.33

qui semble provenir du signal radar. Toutefois, la méthode proposée a su tirer parti de la complémentarité entre les données radar et optique pour améliorer la classification à partir des données multi-sources. C'est également le cas sur le site du Sénégal où la méthode proposée n'est pas la meilleure en employant individuellement les sources mais arrive à être plus performante que les autres approches avec la combinaison des données multi-sources même si toutes les méthodes évaluées se sont améliorées. Il n'existe pas de tendance

réelle dégageant une méthode commune sur les deux sites qui traite mieux que les autres à la fois les données radar ou optique. Néanmoins, il est à souligner, conformément aux résultats obtenus sur les deux sites, que l'approche SVM ne semble pas parfaitement adapté pour exploiter les données radar.

Analyse d'ablation sur les composantes du modèle proposé

Tableau 3.8 – Performances moyennes des modèles

Reunion	F1-score	Kappa	Précision globale
$HOb2sRNN_{NoAtt}$	77.66 ± 0.99	0.749 ± 0.011	77.74 ± 0.99
$HOb2sRNN_{SoftMaxAtt}$	77.32 ± 1.22	0.746 ± 0.013	77.47 ± 1.18
$HOb2sRNN_{NoPreHier}$	78.35 ± 0.70	0.756 ± 0.007	78.43 ± 0.66
$HOb2sRNN_{GRU}$	79.09 ± 0.57	0.764 ± 0.006	79.10 ± 0.50
HOb2sRNN	**79.66** ± 0.85	**0.772** ± 0.009	**79.78** ± 0.82
Sénégal	F1-score	Kappa	Précision globale
$HOb2sRNN_{NoAtt}$	89.86 ± 0.62	0.874 ± 0.008	89.89 ± 0.63
$HOb2sRNN_{SoftMaxAtt}$	89.91 ± 0.54	0.874 ± 0.007	89.92 ± 0.52
$HOb2sRNN_{NoPreHier}$	89.25 ± 0.88	0.866 ± 0.011	89.24 ± 0.87
$HOb2sRNN_{GRU}$	89.12 ± 0.64	0.864 ± 0.008	89.11 ± 0.64
HOb2sRNN	**90.78** ± 1.03	**0.885** ± 0.013	**90.78** ± 1.03

Les résultats de l'étude d'ablation effectuée sur les différentes composantes du modèle *HOb2sRNN* sont présentés dans le tableau 3.8. En ce qui concerne l'utilisation ou non de mécanismes d'attention à travers les différentes variantes ou le modèle proposé, nous pouvons observer que cette composante contribue aux performances finales de classification sur les deux sites d'étude, davantage sur l'île de la Réunion que sur le site du Sénégal (environ 2 points d'amélioration apportés par le modèle proposé par rapport à $HOb2sRNN_{NoAtt}$ contre presque un point). De plus, il est à noter que la variante $HOb2sRNN_{SoftMaxAtt}$présente des performances similaires à celles de la variante $HOb2sRNN_{NoAtt}$ et inférieures à celles du modèle proposé, ce qui confirme notre hypothèse de départ sur le bénéfice qu'induirait la relaxation de la contrainte de somme équivalente à 1 existant sur les poids d'attention. À l'égard de l'évaluation de la stratégie de pré-entraînement du modèle, nous notons également la plus-value de cette étape sur les deux sites d'étude (environ 1 point d'amélioration apporté par le modèle proposé par rapport à $HOb2sRNN_{NoPreHier}$). Ces résultats montrent que l'injection de connaissances spécifiques au domaine à travers un processus de pré-entraînement des réseaux de neurones peut améliorer les performances finales de classification. Enfin, les résultats de la variante $HOb2sRNN_{GRU}$ montrent que l'étape

d'enrichissement réalisée dans la cellule FCGRU se révèle aussi utile sur les deux sites, en particulier quand il existe peu de données étiquetés comme c'est le cas pour le site du Sénégal.

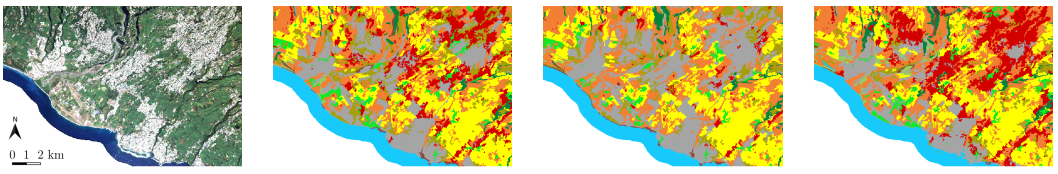
3.2.4 Évaluation qualitative

Extraits des cartes d'occupation du sol

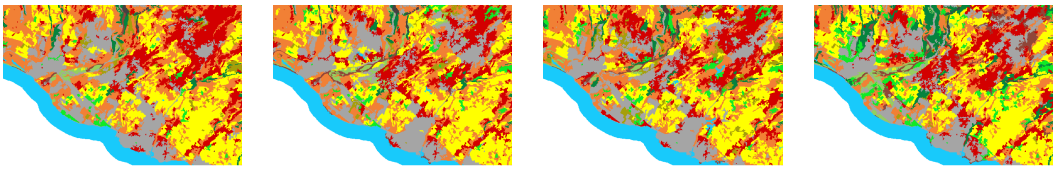
Dans cette première partie de notre évaluation qualitative, nous nous intéressons aux cartes d'occupation du sol produites par le bais des méthodes évaluées. Pour chacun des sites d'étude, nous présentons successivement deux séries d'extraits des cartes générées. Rappelons que les cartes sont produites en labélisant la totalité des objets issus des couches de segmentation respectives des deux sites. Les extraits des cartes sont présentés respectivement dans les figures 3.15 et 3.16. Sur l'île de la Réunion, la première série d'extraits (Figure 3.15 a-h) représente une partie de Saint-Pierre, une zone côtière mixte entre espace urbain et activités agricoles. Dans cet exemple, nous notons les confusions mises en évidence dans l'analyse par classe entre les espaces artificialisés et les cultures sous serre. Visuellement, les modèles RF (RF_{late} en particulier) classifient mieux les espaces artificialisés. La seconde série d'extraits (Figure 3.15 i-p) représente une zone mixte entre agriculture et végétation naturelle environnante au sein du cirque de Salazie. Ici, il est à noter que la méthode proposée détecte une quantité plus réaliste d'arboriculture par rapport aux autres approches. De même, à droite de quelques extraits, nous observons que les espaces boisés sont classifiés à tort par certaines approches comme RF en canne à sucre et landes/savanes. L'extrait de l'approche OD2RNN montre également une confusion entre rochers et eau en bas à gauche de l'extrait.

Sur le site du Sénégal, la première série d'extraits (Figure 3.16 a-h) se situe dans un paysage rural avec une zone humide, près du village de Diohine au sud ouest du site d'étude. Si l'ensemble des méthodes parviennent à bien ressortir le village en question, ce n'est pas le cas pour la zone humide qui est mal détectée particulièrement par les approches RF. Dans le second extrait (Figure 3.16 i-p) mettant également en évidence un paysage rural avec cette fois des activités agricoles, nous remarquons également que les modèles RF ont tendance à surestimer la classe légumineuse tandis que les autres approches détectent la classe céréale ou de la jachère. De cette analyse qualitative sur les cartes d'occupation du sol, il ressort surtout que les prédictions de l'algorithme RF, parfois sensible au déséquilibre entre classes (Maxwell et al., 2018), soient souvent biaisés, plus que celles des autres approches, vers les classes majoritaires dans les jeux de données comme la canne à sucre sur

Extrait 1 en zone côtière mixte urbain et agriculture

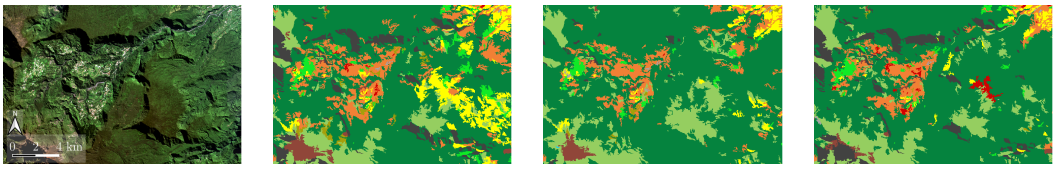


(a) Image RGB (b) RF_{early} (c) RF_{late} (d) SVM

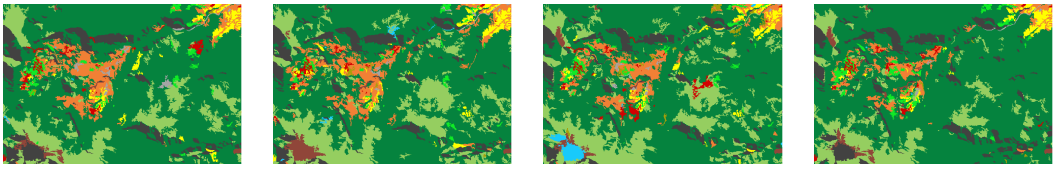


(e) MLP (f) TempCNN (g) OD2RNN (h) *HOb2sRNN*

Extrait 2 en zone mixte agriculture et végétation naturelle environnante



(i) Image RGB (j) RF_{early} (k) RF_{late} (l) SVM



(m) MLP (n) TempCNN (o) OD2RNN (p) *HOb2sRNN*

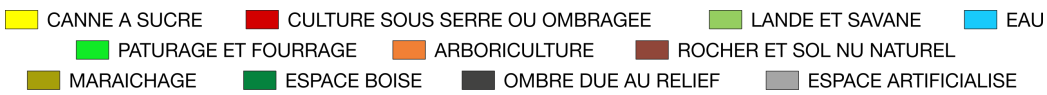
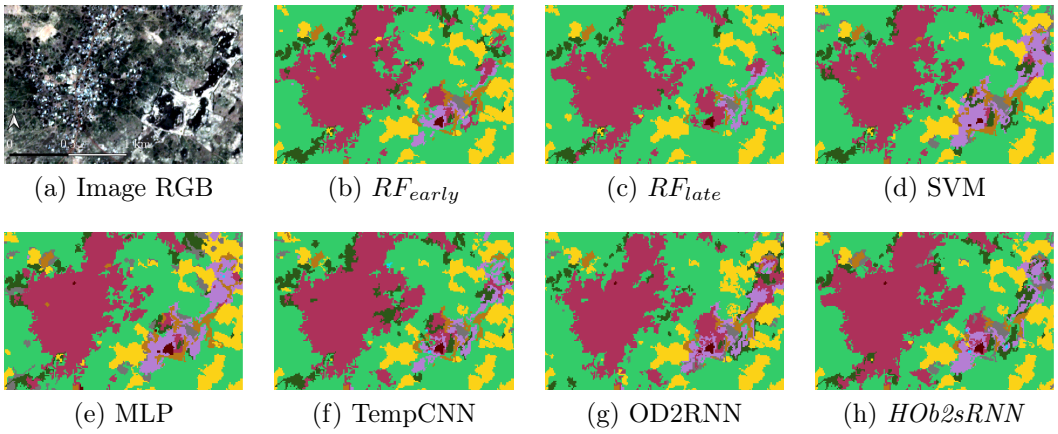


Figure 3.15 – Extraits des cartes d'occupation du sol sur le site de la Réunion. L'image THRS ayant servi à la segmentation est affichée en guise de référence.

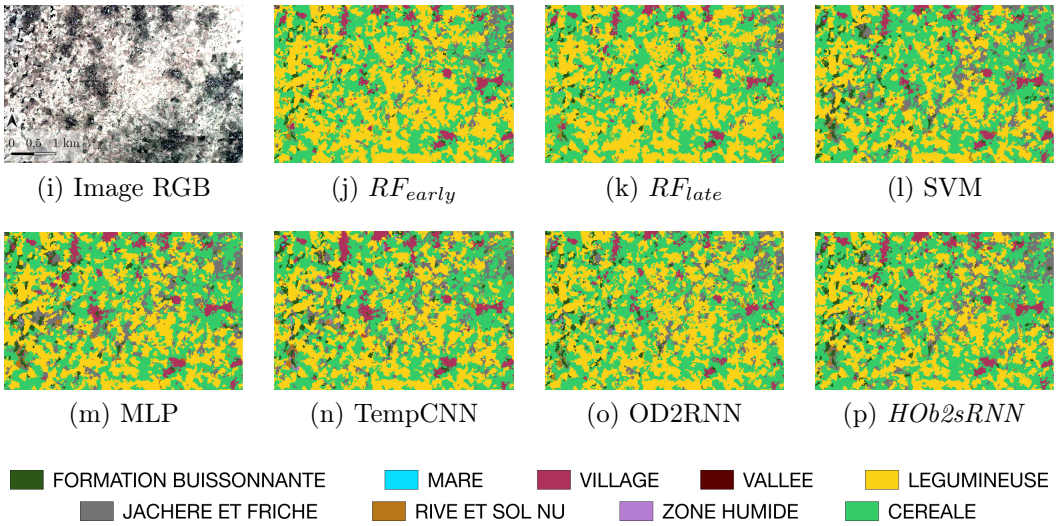
Extrait 1 en paysage rural avec une zone humide



(a) Image RGB (b) RF_{early} (c) RF_{late} (d) SVM

(e) MLP (f) TempCNN (g) OD2RNN (h) *HOb2sRNN*

Extrait 2 en paysage rural avec des activités agricoles



(i) Image RGB (j) RF_{early} (k) RF_{late} (l) SVM

(m) MLP (n) TempCNN (o) OD2RNN (p) *HOb2sRNN*

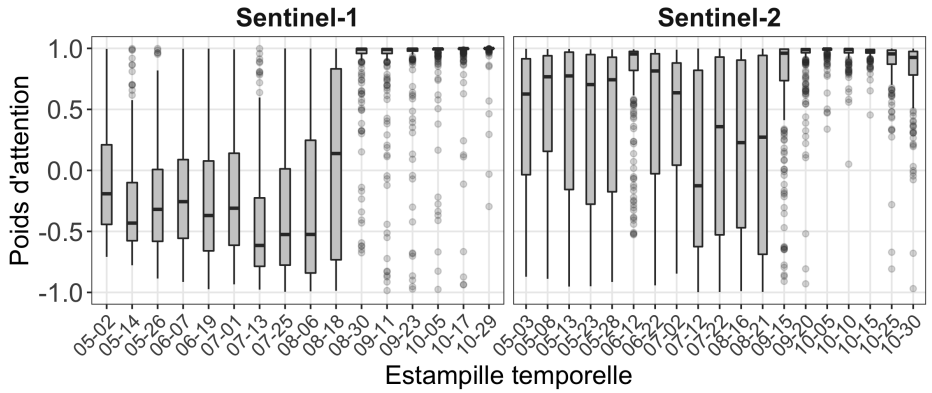
Figure 3.16 – Extraits des cartes d'occupation du sol sur le site du Sénégal. L'image THRS ayant servi à la segmentation est affichée en guise de référence.

la Réunion ou les village et légumineuse dans le cas du site du Sénégal.

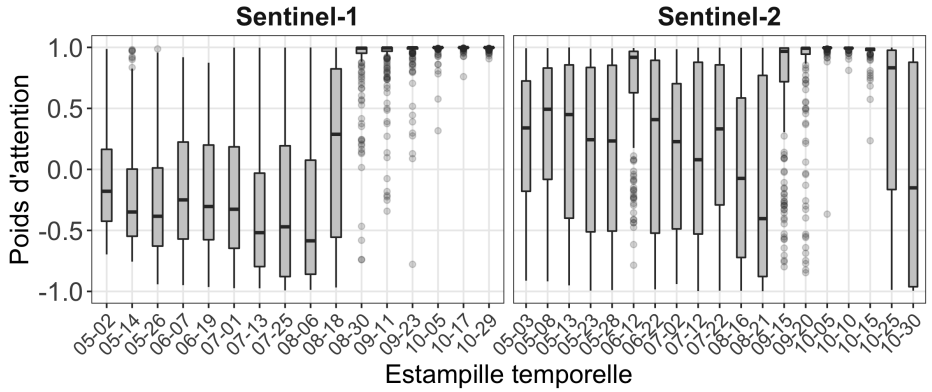
Analyse sur les poids d'attention

Dans cette seconde et dernière partie de l'évaluation qualitative, nous avons examiné dans quelle mesure les informations annexes dérivées des mécanismes d'attention c'est-à-dire les poids d'attention, pourraient contribuer à obtenir des indications significatives sur le comportement de la méthode *HOb2sRNN* vis-à-vis de la tâche de classification de l'occupation du sol. Les poids d'attention ont été employés avec succès dans le domaine du traitement automatique du langage pour expliquer quelles parties du signal d'entrée contribuaient aux décisions prises par les RNNs (Bahdanau et al., 2015; Britz et al., 2017; Choi et al., 2018). Dans le but de mettre en place une analyse analogue dans notre contexte traitant de la classification de séries temporelles multi-sources, nous nous sommes penchés sur les poids d'attention du modèle extraits du jeu test sur le site du Sénégal avec un intérêt particulier pour les cultures (classes céréale et légumineuse) motivé par les connaissances agronomiques que nous avons du site d'étude. Nous reportons sur la figure 3.17, la distribution des poids d'attention obtenus respectivement pour les classes céréale et légumineuse. Par souci de simplicité, nous ne présentons que les résultats issus du mécanisme d'attention associé à la classification des caractéristiques fusionnées.

À première vue, nous observons que le modèle pondère assez similairement sur les deux classes, les estampilles temporelles des différentes sources. Nous remarquons également que les dernières estampilles des séries temporelles sont fortement pondérées. Il vaut la peine de noter qu'il peut y avoir une corrélation entre ces valeurs d'attention élevées et le cycle de croissance de ces cultures. Dans le bassin arachidier sénégalais, ces dernières atteignent leur pic d'activité pendant le mois d'août (milieu de la série temporelle) également caractérisé par une forte activité pluvieuse, ce qui induit plus de discrimination dans les réponses spectrales des classes d'occupation du sol. Néanmoins, sur la distribution des poids présentée, un détail fait la différence au niveau des deux dernières estampilles optiques (25-10-2018 et 30-10-2018) des deux classes qui sont pondérées différemment. Les poids attribués vis-à-vis de ces deux estampilles sont plus élevés pour la classe céréale que pour la classe légumineuse. Ce comportement semble être associé aux pratiques agricoles adoptées sur le site en fin de saison lors de la récolte. Tandis que les céréales (principalement le mil) sont récoltées en coupant uniquement les épis, les légumineuses (principalement l'arachide) sont complètement arrachées de terre. Ainsi, sur les estampilles considérées, les parcelles de céréale sont certainement couvertes de plantes sénescentes alors que les parcelles



(a) Classe Céréale

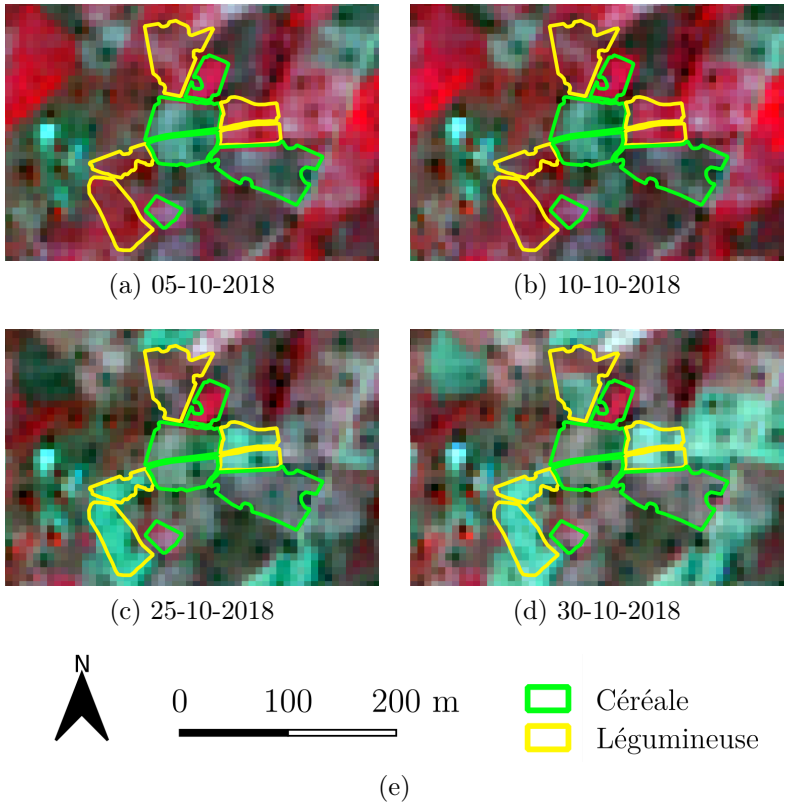


(b) Classe Légumineuse

Figure 3.17 – Distribution des poids d'attention des estampilles temporelles multi-sources pour les classes céréale et légumineuse

de légumineuse se transforment en sol nu. Ces pratiques sont visibles sur les images acquises aux dernières estampilles temporelles. Elles sont illustrées sur la figure 3.18 où nous pouvons observer que la réponse spectrale des parcelles de légumineuse n'est plus la même à partir de l'estampille du 10-10-2018.

En conclusion de cette analyse, nous avons observé des corrélations existantes entre les poids d'attention dérivés du modèle HOb2sRNN et les pratiques agricoles qui caractérisent le site d'étude considéré. Comme le montrent ces résultats, l'exploration des paramètres d'attention peut aider à mieux appréhender certaines décisions prises par le modèle proposé et fournir ainsi des indications précieuses sur l'importance des informations contenues dans les séries temporelles d'images satellitaires.



(a) 05-10-2018 (b) 10-10-2018

(c) 25-10-2018 (d) 30-10-2018

(e)

Figure 3.18 – Visualisation des pratiques agricoles de fin de saison sur le site du Sénégal. Les images Sentinel-2 de fond sont affichées en composition colorée « fausse couleur ».

3.3 Approche MMCNN$_{SD}$

Cette section est consacrée à la méthode *MMCNN$_{SD}$* (Multi-Modal CNN with per source Self-Distillation) qui traite de la cartographie de l'occupation du sol au niveau pixel à partir de données multi-modales ou multi-sources de télédétection en considérant simultanément des séries temporelles radar (Sentinel-1) et optique (Sentinel-2) ainsi qu'une image optique (SPOT) à très haute résolution spatiale. Nous détaillons tout d'abord le fonctionnement de la méthode proposée et présentons ensuite le protocole expérimental ainsi que les résultats de son évaluation.

3.3.1 Description de la méthode

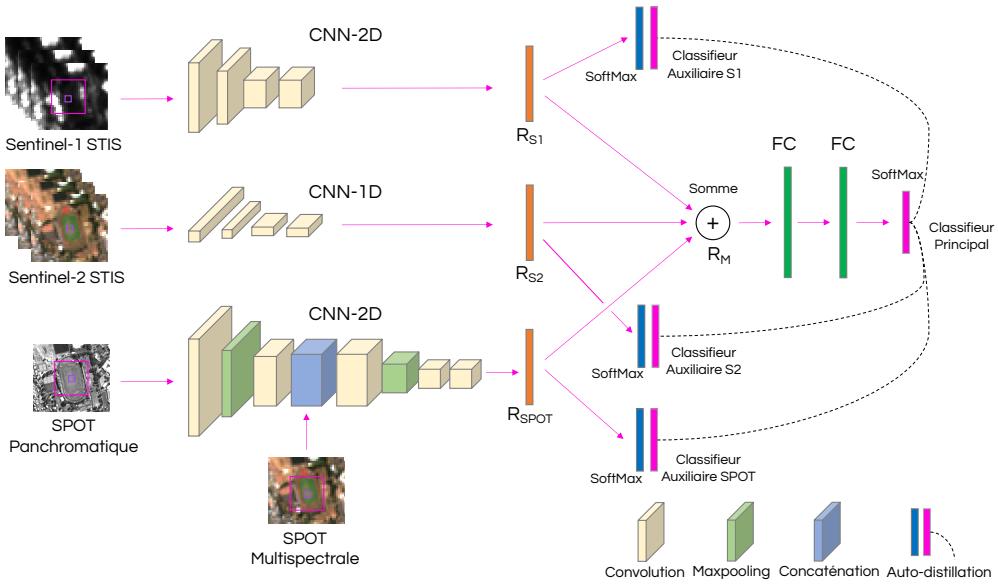


Figure 3.19 – Aperçu global de la méthode *MMCNN$_{SD}$*. Elle a 3 branches, une dédiée à chaque source de données. Les séries temporelles Sentinel-1 (S1 STIS) et l'image SPOT sont traitées par un réseau CNN-2D tandis que les séries temporelles Sentinel-2 (S2 STIS) le sont par un réseau CNN-1D. Les représentations extraites pour chaque source sont ensuite agrégées par la somme pour effectuer la classification finale. À cet effet, sont employés, un classifieur principal associée à la représentation agrégée et des classifieurs auxiliaires par source supervisés à partir de la distillation de la sortie du classifieur principal.

La Figure 3.19 donne un aperçu global du fonctionnement de la méthode

$MMCNN_{SD}$. Elle a 3 branches, une dédiée à chaque source d'information complémentaire qu'elle prend en entrée : les séries temporelles Sentinel-1 et Sentinel-2 ainsi que l'image à très haute résolution spatiale SPOT. Chaque branche est composée d'un encodeur CNN qui extrait une représentation spécifique par source. Les représentations par source, respectivement R_{S1}, R_{S2} et R_{SPOT}, sont par la suite agrégées selon un schéma de fusion tardive (Hu et al., 2019) par le biais de leur somme. La représentation multi-source R_M, résultant de l'agrégation, est finalement transmise à deux couches entièrement connectées suivies par une couche de sortie avec une activation $SoftMax$ (classifieur principal), afin d'effectuer la classification finale. De plus, la méthode $MMCNN_{SD}$ est dotée d'une stratégie d'auto-distillation (Gou et al., 2020; Wang and Yoon, 2021) lui permettant « d'apprendre d'elle-même ». En effet, chaque encodeur CNN par source est également muni d'une couche de sortie (classifieur auxiliaire) avec une activation $SoftMax$, dans le but de forcer le modèle à extraire des représentations plus discriminantes et complémentaires entre elles. Les classifieurs auxiliaires par source sont entraînés pour mimer la sortie du classifieur principal, ceci dans le but de distiller ou transférer les connaissances des couches les plus profondes, notamment la sortie du modèle, vers les couches les moins profondes c'est-à-dire celles des encodeurs par source. Tandis que le processus classique de distillation de connaissances (Hinton et al., 2015; Wang and Yoon, 2021) est basé sur une architecture « enseignant – étudiant » où le transfert de connaissances se fait de l'enseignant à l'étudiant, celui de l'auto-distillation (Zhang et al., 2019) ne requiert pas de paire de modèles distincts puisque la distillation de connaissances se fait à partir du modèle lui même de manière autonome. Pour faire le lien avec l'architecture « enseignant – étudiant », dans notre cas, la sortie du classifieur principal du modèle $MMCNN_{SD}$ peut être assimilée aux connaissances du modèle enseignant tandis que les encodeurs par source représentent les modèles étudiants ayant pour objectif de mimer le comportement de l'enseignant. Nous employons de cette façon la stratégie d'auto-distillation dans un contexte d'analyse multi-modale. À cette fin, la perte (L) à minimiser dans l'entraînement du modèle est définie comme suit :

$$L = CE(Y, CL(R_M)) + \lambda \sum_{s \in \{S1, S2, SPOT\}} CE(CL(R_M), OUT(R_s)) \tag{3.7}$$

où Y est l'information de référence c'est-à-dire la classe d'occupation du sol à prédire ; CE est l'entropie croisée dans un contexte multi-classe ; CL correspond au classifieur principal c'est-à-dire un réseau de neurones constitué de deux couches entièrement connectées avec une activation ReLU et la normalisation par lot, suivies d'une couche de sortie avec une activation $SoftMax$;

OUT correspond à un classifieur auxiliaire constitué d'une couche de sortie avec une activation $SoftMax$. Enfin, λ est un hyper-paramètre contrôlant l'importance relative des coûts liés aux classifieurs auxiliaires et donc à la stratégie d'auto-distillation par rapport à celle du classifieur principal associé à la représentation multi-modale. Bien que le modèle implique, pendant la phase d'entraînement, l'utilisation du classifieur principal et des classifieurs par source associés à la stratégie d'auto-distillation, seule la prédiction fournie par le classifieur principal, c'est-à-dire $CL(R_M)$, est considérée en temps d'inférence. L'ensemble des paramètres associés au modèle $MMCNN_{SD}$ est appris de bout en bout.

Architecture des encodeurs CNN par source

Afin de tirer partie de la complémentarité des diverses sources d'information dont nous disposons, nous avons conçu des encodeurs CNN qui leur sont spécifiques.

Pour traiter la série temporelle Sentinel-1, nous utilisons un réseau convolutif à deux dimensions (CNN-2D). Ainsi, les données Sentinel-1 sont organisés en un empilement d'images successives dont le nombre total de bandes est équivalent au nombre de dates d'acquisitions multiplié par un facteur de 2 puisque les données Sentinel-1 sont analysés en double polarisation (VV et VH). Des imagettes de mêmes dimensions spatiales centrées sur le pixel à classifier sont dès lors extraites de cet empilement et constituent l'information en entrée du CNN-2D.

Pour traiter la série temporelle Sentinel-2, nous avons suivi la littérature récente en cartographie de l'occupation du sol (Pelletier et al., 2019) et adopté un réseau convolutif unidimensionnel (CNN-1D). Ici, c'est l'information séquentielle du pixel, provenant de la série temporelle, qui constitue l'entrée du modèle CNN-1D.

Enfin pour traiter l'image à très haute résolution spatiale SPOT, nous adoptons à nouveau un CNN-2D, à fortiori dans le but d'exploiter, autant que possible, cette information spatiale à échelle fine. Les images SPOT disposent d'une bande panchromatique et de bandes multispectrales à des résolutions spatiales différentes (1.5-m et 6-m respectivement). Afin de traiter les deux types d'informations à leur résolution native en évitant dans le mesure du possible toute étape intermédiaire pouvant engendrer des coûts de calcul supplémentaires comme le rééchantillonnage ou le pan-sharpening (Gaetano et al., 2018), le modèle CNN-2D adopté traite en premier l'information panchromatique et une fois que les cartes d'activation produites ont les mêmes dimensions spatiales que les bandes multispectrales, ces dernières sont intégrées au processus par concaténation. Ici également, des imagettes (pan-

chromatique et multispectrales), prises autour de la position géographique correspondante au centre du pixel à classifier sur l'image Sentinel, sont extraites de l'image SPOT.

Tableau 3.9 – Détails sur l'architecture du modèle $MMCNN_{SD}$. (Par souci de lisibilité, les classifieurs auxiliaires sont omis.)

Sentinel-1	**Sentinel-2**	**SPOT**
		7×7 Conv2D (128) avec PAN
		MaxPooling2D 3×3
3×3 Conv2D (128)	5×1 Conv1D (128)	5×5 Conv2D (256)
3×3 Conv2D (128)	3×1 Conv1D (128)	Concaténation avec MS
3×3 Conv2D (256)	3×1 Conv1D (256)	3×3 Conv2D (256)
1×1 Conv2D (256)	1×1 Conv1D (256)	MaxPooling2D 3×3
GlobAvgPooling2D	GlobAvgPooling1D	3×3 Conv2D (256)
		1×1 Conv2D (256)
		GlobAvgPooling2D
Agrégation par somme		
Couche entièrement connectée (512) + ReLU + Normalisation par lot		
Couche entièrement connectée (512) + ReLU + Normalisation par lot		
Couche de sortie entièrement connectée avec activation $SoftMax$		

Nous reportons dans le tableau 3.9, les détails concernant l'architecture du modèle $MMCNN_{SD}$. La partie initiale du tableau, incluant les couches de Pooling, décrit les encodeurs CNN par source précédemment discutés. Conv1D et Conv2D désignent respectivement des convolutions 1D et 2D. Les valeurs associées (128, 256 et 512) correspondent au nombre de filtres de convolution. Chaque couche convolutive est activée par une fonction ReLU, suivie de la normalisation par lot et d'une couche de Dropout.

3.3.2 Protocole expérimental

L'évaluation du modèle $MMCNN_{SD}$ est conduite sur les sites de la Réunion et de la Dordogne. Le site du bassin arachidier au Sénégal n'a pas été considéré dans cette partie par souci d'uniformisation du modèle, en l'occurrence l'encodeur CNN de l'image à très haute résolution spatiale. Nous rappelons que les images PlanetScope disponibles sur le site du Sénégal ne disposent pas de bande panchromatique comme les images SPOT mais que de bandes multispectrales.

Les dates d'acquisition des séries temporelles Sentinel-1 et Sentinel-2 sont illustrées par les figures 3.4 et 3.20, respectivement pour l'île de la Réunion et le site de la Dordogne. L'image SPOT acquise sans nuages sur le site la Dordogne date du 03 Mars 2016 tandis que celle de l'île de la Réunion résulte

d'un mosaïquage par une technique d'harmonisation colorimétrique (Cresson and Saint-Geours, 2015) entre 4 images SPOT, acquises respectivement les 26 Décembre 2016, 10 Mai, 11 Juin et 20 Novembre 2017, dans le but d'assurer une couverture à très haute résolution sans nuages de l'île. Dans cette évaluation également, nous soulignons que seules les bandes spectrales à 10-m sont considérées pour les données Sentinel-2, en plus des indices de végétation NDVI et NDWI.

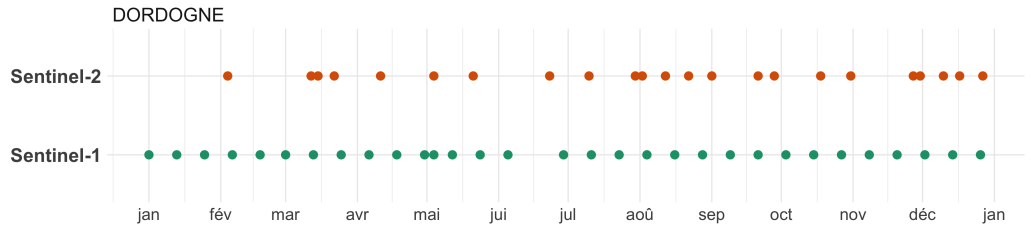


Figure 3.20 – Dates d'acquisition des 31 images Sentinel-1 et 23 images Sentinel-2 sur le site de la Dordogne

La vérité terrain disponible en format vecteur sur les deux sites d'étude (voir chapitre 2) a été rastérisé à la résolution spatiale des images Sentinel (10-m) obtenant un total de plus de 800 000 pixels labélisés sur chacun des sites (Tableaux 3.10 et 3.11).

Tableau 3.10 – Nombre de pixels par classe sur l'île de la Réunion

Classe	Pixels
1 – CANNE À SUCRE	88 962
2 – PÂTURAGE ET FOURRAGE	68 098
3 – MARAÎCHAGE	17 488
4 – CULTURE SOUS SERRE OU OMBRAGÉE	1 908
5 – ARBORICULTURE	33 721
6 – ESPACE BOISE	205 023
7 – LANDE ET SAVANE	155 231
8 – ROCHER ET SOL NU NATUREL	154 343
9 – OMBRE DUE AU RELIEF	54 301
10 – EAU	82 592
11 – ESPACE ARTIFICIALISÉ	19 056
TOTAL	880 723

Dans le processus d'évaluation du modèle, nous avons tout d'abord réalisé une étude spécifique sur les encodeurs CNN par source afin de valider nos choix architecturaux. Pour ce faire, les données Sentinel-1 et Sentinel-2 sont

Tableau 3.11 – Nombre de pixels par classe sur le site de la Dordogne

Classe	Pixels
1 – ESPACE ARTIFICIALISÉ	2 002
2 – EAU	50 471
3 – FORÊT	378 969
4 – LANDES	99 627
5 – ARBORICULTURE	97 546
6 – VIGNE	92 259
7 – AUTRES CULTURES	93 562
TOTAL	814 436

analysées en considérant successivement des réseaux CNN-1D, CNN-2D et CNN-3D. Les CNN-1D/2D sont les mêmes que ceux adoptés pour le modèle (voir Tableau 3.9). En ce qui concerne le CNN-3D, nous avons adopté un modèle similaire aux deux premiers en gardant presque la même architecture notamment le même nombre de couches convolutives et de filtres associés ainsi qu'une couche de Pooling pour l'extraction des caractéristiques. Seuls la taille des filtres de convolution et le pas dans le domaine temporel sont modifiés pour les couches convolutives. Ainsi, une taille de $3 \times 3 \times 3$ est adoptée pour les trois premières couches convolutives suivant les recommandations de Ji et al. (2018) tandis que nous avons gardé une taille de 1 dans les 3 dimensions pour la dernière couche convolutive, similairement aux CNN-1D/2D. Par ailleurs, le pas des convolutions dans le domaine temporel est fixé à 2 pour les deuxième et troisième couches convolutives afin de tirer d'avantage parti du signal temporel. L'architecture du modèle CNN-3D est reportée dans le tableau 3.12. Notons que le module de classification des encodeurs CNN-1D/2D/3D est équivalent à celui du modèle $MMCNN_{SD}$ c'est-à-dire son classifieur principal.

Tableau 3.12 – Détails sur l'architecture de l'encodeur CNN-3D.

Encodeur CNN-3D
$3 \times 3 \times 3$ Conv3D (128) avec pas (1,1,1)
$3 \times 3 \times 3$ Conv3D (128) avec pas (1,1,2)
$3 \times 3 \times 3$ Conv3D (256) avec pas (1,1,2)
$1 \times 1 \times 1$ Conv3D (256) avec pas (1,1,1)
GlobalAvgPooling3D

Par la suite, nous avons intégré dans l'analyse la combinaison des données multi-modales par le biais de la méthode $MMCNN_{SD}$. À cette fin, nous avons réalisé également une étude d'ablation sur la combinaison des sources afin de

démêler leurs interactions. Deux variantes du modèle proposé avec Sentinel-1 et Sentinel-2 uniquement puis Sentinel-2 et SPOT ont ainsi été testées. Ces variantes sont dénommées : $MMCNN_{SD}^{S1+S2}$ et $MMCNN_{SD}^{S2+SPOT}$. De plus, nous évaluons à travers deux autres variantes la stratégie d'auto-distillation proposée et la contribution des classifieurs auxiliaires par source qui y sont associés afin de comprendre leur bénéfice par rapport à notre modèle. La première variante dénommée $MMCNN_{noSD}$, peut être assimilée à un processus de fusion tardive classique sans supervision auxiliaire tel que présentée par Hong et al. (2020) et sans auto-distillation. La seconde variante désignée par $MMCNN_{HardLabels}$ s'appuie sur les travaux récents comme Benedetti et al. (2018) ou Interdonato et al. (2019) ayant employé dans leur processus de fusion tardive une supervision auxiliaire où des classifieurs distincts par source sont entraînés à partir des étiquettes originales. Par ailleurs, nous considérons dans cette évaluation une adaptation directe du modèle $M^3Fusion$ proposé récemment Benedetti et al. (2018). Le modèle $M^3Fusion$ a été mis au point pour réaliser une cartographie multi-source de l'occupation du sol à partir de séries temporelles Sentinel-2 et d'une image THRS SPOT. Il traite les données en entrée au moyen de deux branches spécifiques, l'une basée sur un RNN pour encoder les séquences Sentinel-2 et l'autre sur un CNN-2D pour traiter les imagettes SPOT. Afin d'effectuer une comparaison équitable dans notre cas, nous avons équipé le modèle originel d'une branche RNN additionnelle afin de traiter les données Sentinel-1. Cette adaptation du modèle $M^3Fusion$ peut être également perçue comme une variante du modèle $HOb2sRNN$ étudié dans la section précédente à qui l'on aurait rajouté une branche CNN pour encoder l'information spatiale fine. Cette variante est néanmoins dépourvue de la modification du mécanisme d'attention et de la stratégie de pré-entraînement tenant compte des relations de hiérarchie entre classes d'occupation du sol. Pour finir, nous avons étudié l'effet de la variation de la dimensionnalité des caractéristiques extraites par source ainsi que celui de la variation de l'hyper-paramètre λ contrôlant la stratégie d'auto-distillation.

En ce qui concerne les données en entrée des modèles, comme évoqué auparavant, nous avons extrait des imagettes pour décrire un emplacement géographique spécifique. Ainsi, la taille des imagettes Sentinel-1/2 est fixée à 9×9 soit 4 pixels dans chaque direction autour du pixel cible à classifier tandis que pour les images SPOT, des tailles de 8×8 et 32×32 sont prises respectivement pour les bandes multispectrales et panchromatique, pareillement à Benedetti et al. (2018). Rappelons que les imagettes SPOT sont extraites autour du même emplacement géographique correspondant au centre du pixel à classifier sur les images Sentinel. Rappelons également que l'encodeur CNN-1D ne prend uniquement en compte que l'information sé-

quentielle provenant du pixel central des imagettes extraites. Par ailleurs, afin de répondre aux exigences en termes de données en entrée du modèle $M^3Fusion$, un processus de pan-sharpening a été réalisé entre la bande panchromatique et les bandes multi-spectrales de l'image SPOT afin d'extraire des imagettes multi-spectrales de taille 32×32 à la résolution spatiale de 1.5-m.

Pour le reste, les valeurs des imagettes sont normalisées par bande, considérant les série temporelles Sentinel ou l'image SPOT, dans l'intervalle [0,1]. Les données sont divisées en jeu d'entraînement, jeu de validation et jeu test avec des proportions respectives de 50%, 20% et 30%. Pour ce faire, nous nous assurons aussi que tous les pixels appartenant au même polygone échantillonné (portant le même identifiant) se retrouvent exclusivement dans l'une des 3 partitions (entraînement, validation ou test) afin d'éviter au mieux de possibles biais d'auto-corrélation spatiale dans la procédure d'évaluation. Par ailleurs, les modèles sont optimisés à travers une procédure d'entraînement et validation, consistant à sauvegarder les paramètres ou poids permettant à un modèle au moment de l'entraînement de généraliser le mieux sur le jeu de validation.
La phase d'entraînement est conduite sur 300 époques pour l'ensemble des modèles en utilisant l'optimiseur Adam (Kingma and Ba, 2015) avec un taux d'apprentissage de 10^{-4}. Le taux de Dropout est fixé à 0.4 et la taille des lots à 256 échantillons. L'hyper-paramètre λ est quant à lui fixé à 0.3 pour toutes les approches multi-sources incluant des classifieurs auxiliares par source.
Les métriques précision globale, F1-score et cœfficient Kappa sont considérées pour évaluer les performances des modèles sur le jeu test. Étant donné que les performances sont susceptibles de varier en fonction de la division des données, en raison d'échantillons plus simples ou plus complexes impliqués dans les différentes partitions, ces métriques sont moyennées sur 5 divisions aléatoires du jeu de données suivant la stratégie précédemment décrite. Les expériences sont réalisées sur une station de travail dotée d'un CPU AMD Ryzen 7 3700X, d'une RAM de 64 GB et d'une carte GPU NVIDIA RTX 2080. Le nombre de paramètres à optimiser ainsi que les temps de calcul des différents modèles sont présentés dans le tableau 3.13. Les implémentations des modèles sont disponibles à l'adresse `https://github.com/eudesyawog/S1S2VHSR`.

3.3.3 Évaluation quantitative

Encodeurs CNN par source

Nous reportons respectivement dans les tableaux 3.14 et 3.15, les performances moyennes des encodeurs CNN par source sur les deux sites d'étude.

Tableau 3.13 – Nombre de paramètres optimisables des différents modèles et temps de calcul associés sur les 300 époques d'entraînement

Sources et Modèles		Paramètres optimisables		Temps de calcul	
		REUNION	DORDOGNE	REUNION	DORDOGNE
S1	CNN-1D	0.62 M	0.62 M	0.37 h	0.40 h
	CNN-2D	0.97 M	0.97 M	0.61 h	0.58 h
	CNN-3D	1.80 M	1.80 M	7.54 h	8.40 h
S2	CNN-1D	0.62 M	0.62 M	0.38 h	0.37 h
	CNN-2D	1.05 M	1.07 M	0.84 h	0.81 h
	CNN-3D	1.82 M	1.81 M	6.35 h	6.48 h
SPOT		2.48 M	2.48 M	2.32 h	2.19 h
$M^3Fusion$		12.6 M	12.58 M	15.37 h	15.80 h
$MMCNN_{SD}^{S1+S2}$		1.20 M	1.20 M	0.96 h	0.93 h
$MMCNN_{SD}^{S2+SPOT}$		2.71 M	2.70 M	2.71 h	2.53 h
$MMCNN_{SD}$		3.28 M	3.29 M	3.39h	3.18 h

En premier, nous constatons que tirer parti des dépendances spatiales ou temporelles des données Sentinel-1/2 conduit à des résultats différents. En effet, les convolutions 2D (spatiales) sont largement plus efficaces que les convolutions 1D (temporelles) pour les données Sentinel-1 tandis que les performances sont comparables pour les données Sentinel-2. Ce gain en performance observé sur les deux sites par rapport à l'emploi des convolutions 2D pour les données Sentinel-1 peut se traduire par le fait que les filtres convolutifs 2D aident à atténuer encore plus le bruit spatial lié au chatoiement des images SAR (Wang et al., 2017), ce qui permet sans doute d'obtenir des représentations plus discriminantes. En ce qui concerne l'encodeur CNN-3D, ses performances sont légèrement en dessous du CNN-2D (ex. avec Sentinel-1) ou similaires à celles des autres modèles (ex. avec Sentinel-2) sauf dans le cas de l'île de la Réunion avec les données Sentinel-2 où l'on observe un léger gain par rapport au CNN-1D. Toutefois, le bénéfice qu'apporte l'emploi de convolutions à la fois dans les dimensions spatiale et temporelle (convolutions 3D) reste minimal, surtout si l'on prend en compte le nombre de paramètres largement plus important du CNN-3D par rapport aux autres encodeurs ainsi que ses temps en calcul plus longs (voir Tableau 3.13). Notons par ailleurs, l'importance relative de l'information à très haute résolution spatiale fournie par l'image SPOT sur les deux sites d'étude. Si les performances obtenues avec l'image SPOT sur l'île de la Réunion sont compétitives par rapport à celles de toute la série temporelle Sentinel-2, ce n'est pas le cas sur site de la Dordogne. Pour résumer, cette étude spécifique sur les encodeurs CNN par source suggère que le CNN-2D est plus pertinent pour la représentation des

données Sentinel-1 tandis que, pour des raisons de parcimonie (nombre de paramètres plus léger et temps de calcul réduit), il est plus approprié d'encoder la série temporelle Sentinel-2 avec un CNN-1D. Ci-après, les résultats présentés pour Sentinel-1 et Sentinel-2 sont respectivement ceux obtenus avec les encodeurs CNN-2D et CNN-1D.

Tableau 3.14 – Performances moyennes des encodeurs CNN par source sur l'île de la Réunion

Sources		F1-score	Kappa	Précision globale
S1	CNN-1D	64.82 ± 1.32	0.587 ± 0.018	65.63 ± 1.64
	CNN-2D	**73.09** ± 2.62	**0.684** ± 0.030	**73.39** ± 2.66
	CNN-3D	72.35 ± 2.94	0.673 ± 0.036	72.63 ± 3.16
S2	CNN-1D	87.98 ± 1.12	0.859 ± 0.017	88.09 ± 1.06
	CNN-2D	87.41 ± 1.61	0.851 ± 0.021	87.41 ± 1.66
	CNN-3D	**88.62** ± 1.45	**0.866** ± 0.017	**88.66** ± 1.36
SPOT		88.35 ± 1.33	0.862 ± 0.017	88.35 ± 1.39

Tableau 3.15 – Performances moyennes des encodeurs CNN par source sur le site de la Dordogne

Sources		F1-score	Kappa	Précision globale
S1	CNN-1D	73.54 ± 2.96	0.644 ± 0.028	75.15 ± 2.76
	CNN-2D	**80.50** ± 2.17	**0.730** ± 0.024	**80.73** ± 2.21
	CNN-3D	78.87 ± 3.12	0.709 ± 0.034	79.43 ± 2.88
S2	CNN-1D	**85.97** ± 2.15	**0.806** ± 0.025	86.04 ± 2.01
	CNN-2D	85.90 ± 1.92	**0.806** ± 0.018	**86.05** ± 1.66
	CNN-3D	85.29 ± 2.35	0.793 ± 0.024	84.88 ± 2.46
SPOT		81.75 ± 2.53	0.745 ± 0.028	81.39 ± 2.62

Combinaison multi-modale

Les performances moyennes des modèles multi-sources sont présentées respectivement dans les tableaux 3.16 et 3.17 pour les deux sites d'étude. Nous remarquons tout d'abord que la combinaison de données multi-modales améliore les performances moyennes de classification par rapport aux résultats précédents obtenus avec les sources individuelles. La combinaison simultanée de toutes les sources d'information disponibles par le biais du modèle proposé se révèle être la plus efficace, particulièrement sur l'île de la Réunion. Ceci

Tableau 3.16 – Performances moyennes des modèles CNN multi-sources sur l'île de la Réunion

Sources	F1-score	Kappa	Précision globale
$M^3Fusion$	92.58 ± 0.51	0.912 ± 0.006	92.59 ± 0.50
$MMCNN_{SD}^{S1+S2}$	91.99 ± 0.42	0.906 ± 0.004	92.05 ± 0.30
$MMCNN_{SD}^{S2+SPOT}$	93.07 ± 1.18	0.918 ± 0.014	93.12 ± 1.16
$MMCNN_{SD}$	**94.34** ± 0.49	**0.934** ± 0.006	**94.38** ± 0.49
$MMCNN_{noSD}$	93.21 ± 0.79	0.920 ± 0.009	93.25 ± 0.77
$MMCNN_{HardLabels}$	93.74 ± 0.94	0.926 ± 0.011	93.77 ± 0.96

Tableau 3.17 – Performances moyennes des modèles CNN multi-sources sur le site de la Dordogne

Sources	F1-score	Kappa	Précision globale
$M^3Fusion$	87.16 ± 1.47	0.825 ± 0.017	87.48 ± 1.51
$MMCNN_{SD}^{S1+S2}$	87.09 ± 1.86	0.823 ± 0.020	87.33 ± 1.78
$MMCNN_{SD}^{S2+SPOT}$	88.36 ± 1.70	0.840 ± 0.020	88.56 ± 1.62
$MMCNN_{SD}$	**88.73** ± 1.80	**0.845** ± 0.021	**88.90** ± 1.68
$MMCNN_{noSD}$	87.87 ± 1.73	0.832 ± 0.020	87.94 ± 1.54
$MMCNN_{HardLabels}$	88.20 ± 1.72	0.836 ± 0.021	88.18 ± 1.69

démontre à nouveau le potentiel pour la classification de l'occupation du sol de combiner des données multi-modales de télédétection.

Par rapport aux autres aspects étudiés du modèle proposé, notamment l'apport des classifieurs auxiliaires associés à la stratégie d'auto-distillation, nous remarquons dans l'étude d'ablation menée ($MMCNN_{noSD}$ vs $MMCNN_{HardLabels}$ vs $MMCNN_{SD}$) que ces composants architecturaux contribuent aux performances obtenues sur les deux sites d'étude. Tout d'abord, les modèles équipés de classifieurs auxiliaires c'est-à-dire $MMCNN_{HardLabels}$ et $MMCNN_{SD}$ sont plus performants que la variante assimilée à la fusion tardive classique c'est-à-dire $MMCNN_{noSD}$. Pour investiguer plus en détail cette observation en nous focalisant sur l'approche proposée, nous avons illustré sur la figure 3.21, les comportements des modèles $MMCNN_{noSD}$ et $MMCNN_{SD}$ pendant la phase d'apprentissage en considérant leurs performances sur les jeux d'entraînement et de validation. S'il est à noter que les deux modèles arrivent très bien à s'adapter au jeu d'entraînement de façon quasi identique, c'est bien le modèle proposé et entraîné avec la stratégie d'auto-distillation qui performe le mieux sur le jeu de validation. Ceci montre que la stratégie d'auto-distillation contribue à une meilleure généralisation du modèle. En second lieu, la comparaison en termes de performances numériques entre le modèle proposé et celui qui adopte la stratégie de supervision des classifieurs

auxiliaires à partir des étiquettes originales ($MMCNN_{HardLabels}$), montre également que la stratégie d'auto-distillation améliore l'exploitation conjointe des données multi-modales.

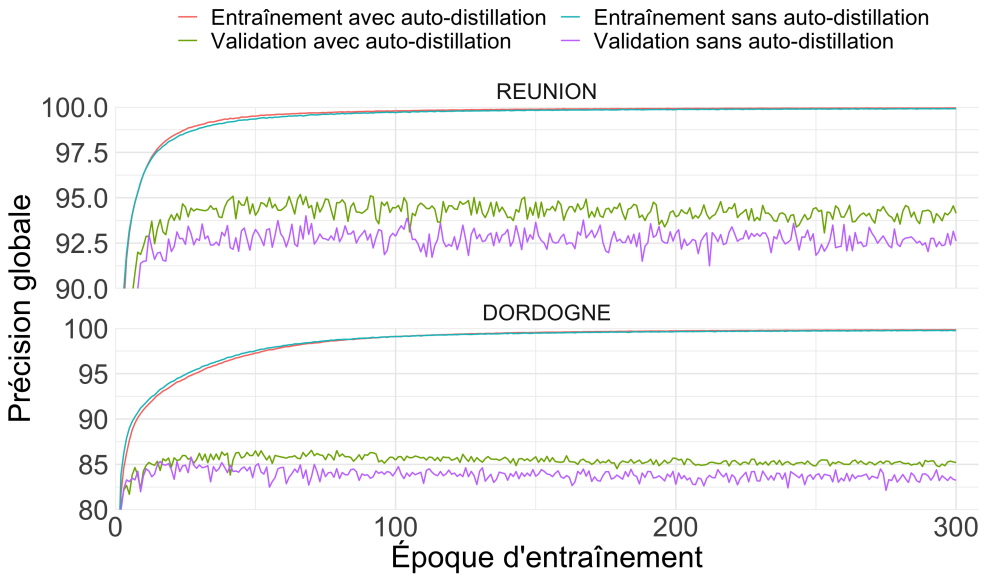


Figure 3.21 – Comportement du modèle avec et sans stratégie d'auto-distillation pendant la phase d'apprentissage

Dans l'ensemble, les résultats obtenus sont conformes aux études récentes sur la distillation des connaissances (Gou et al., 2020; Wang and Yoon, 2021) où il est montré que les étiquettes fournis par le modèle enseignant (dans notre cas le classifieur principal) transmettent plus d'informations utiles et sont plus convenables aux étudiants (dans notre cas les classifieurs auxiliaires) que les étiquettes originales, leur permettant ainsi de mimer plus facilement le comportement du modèle enseignant.

Effet de la variation des hyper-paramètres

Dans cette partie, nous analysons deux hyper-paramètres essentiels du modèle proposé. Nous évaluons comment i) la dimensionnalité des caractéristiques extraites par source et ii) l'hyper-paramètre λ contrôlant la stratégie d'auto-distillation influencent les performances du modèle proposé. Nous faisons varier le premier hyper-paramètre entre les valeurs {64,128,256,512} tandis que le second est évalué entre {0.1,0.2,0.3,0.4,0.5}. Les résultats sont présentés respectivement en figures 3.22 et 3.23.

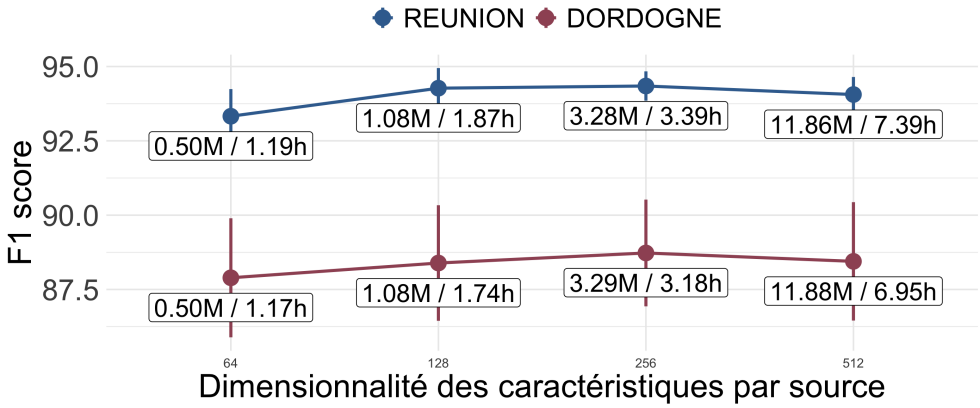


Figure 3.22 – Effet de la variation de la dimensionnalité des caractéristiques par source sur les performances du modèle. L'écart-type est affiché sous forme de barre d'erreur. Les paramètres optimisables ainsi que les coûts en temps de calcul sont indiqués à côté.

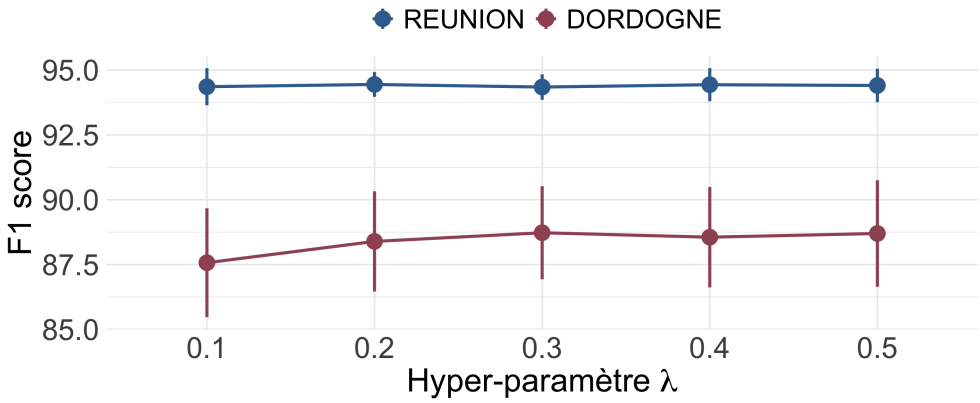


Figure 3.23 – Effet de la variation de l'hyper-paramètre λ contrôlant la stratégie d'auto-distillation sur les performances du modèle. L'écart-type est affiché sous forme de barre d'erreur.

L'analyse sur la dimensionnalité des caractéristiques par source nous montre que des représentations avec 256 caractéristiques chacune semblent convenable pour la méthode proposée. Sur les deux sites d'étude, les performances sont relativement stables (entre 93% et 94% de F1-score sur la Réunion et autour de 88% sur la Dordogne) sur la gamme de valeurs considérée. Plus encore, il convient de noter que le modèle peut déjà bien généraliser avec seulement des représentations de 64 caractéristiques par source, ce qui contribuerait à réduire significativement le nombre de paramètres opti-

misables ainsi que le coût en temps de calcul.

En ce qui concerne l'analyse sur l'hyper-paramètre λ, ici également nous obtenons des performances plus ou moins stables sur les deux sites d'étude pour des valeurs supérieures à 0.2. Il convient de noter que la variation de l'hyper-paramètre λ sur la gamme de valeurs prises en compte, n'influence pas véritablement le bon déroulé de la stratégie d'auto-distillation et par conséquent le comportement du modèle.

Analyse sur les classes d'occupation du sol

La suite de notre évaluation est portée sur l'analyse des performances par classe d'occupation du sol. Les valeurs de F1-score par classe sur les deux sites sont illustrés respectivement par les figures 3.24 et 3.25. Nous observons que la combinaison de sources complémentaires entre elles est pleinement bénéfique pour la plupart des classes d'occupation du sol. Quelques exemples saillants sur l'île de la Réunion en guise d'illustration sont les classes des cultures sous serre ou ombragées, de maraîchage, d'arboriculture ou encore des espaces artificialisés. Le F1-score des cultures sous serre ou ombragées en l'occurrence s'est amélioré de 50% avec Sentinel-2 uniquement à 75% avec le modèle proposé. La caractérisation de cette classe d'occupation a particulièrement bénéficié de l'apport de l'information à très haute résolution spatiale fournie par l'image SPOT (presque 67% de F1-score à elle seule). Le bénéfice est similaire pour les espaces artificialisés ainsi que l'arboriculture qui sont mieux discriminés avec la fine information spatiale. Sur le site de la Dordogne également, nous pouvons souligner comme classes tirant tout particulièrement avantage de la combinaison multi-modale, les espaces artificialisés ainsi que les classes des cultures.

Nous avons analysé également les erreurs commises par les modèles dans la caractérisation des classes d'occupation du sol. Les matrices de confusion sur les deux sites sont reportées respectivement sur les figures 3.26 et 3.27. La tendance observée dans l'analyse des scores par classe se confirme avec les matrices de confusion. Plus il y a de sources complémentaires qui sont combinées, moins il reste de confusions entre les classes d'occupation du sol. Seules quelques erreurs de classification mineures subsistent sur l'île de la Réunion avec le modèle proposé, notamment celles entre les cultures sous serre et ombragées et les espaces artificialisés. Sur le site de la Dordogne, les confusions majeures entre les classes landes et forêt sont également atténuées. Dans l'ensemble, la combinaison simultanée d'informations multi-capteurs, multi-temporelles et multi-échelles s'est avérée précieuse pour mieux caractériser les classes d'occupation du sol présentant non seulement des dépendances temporelles, telles que celles liées aux cultures ou à la végétation naturelle, mais

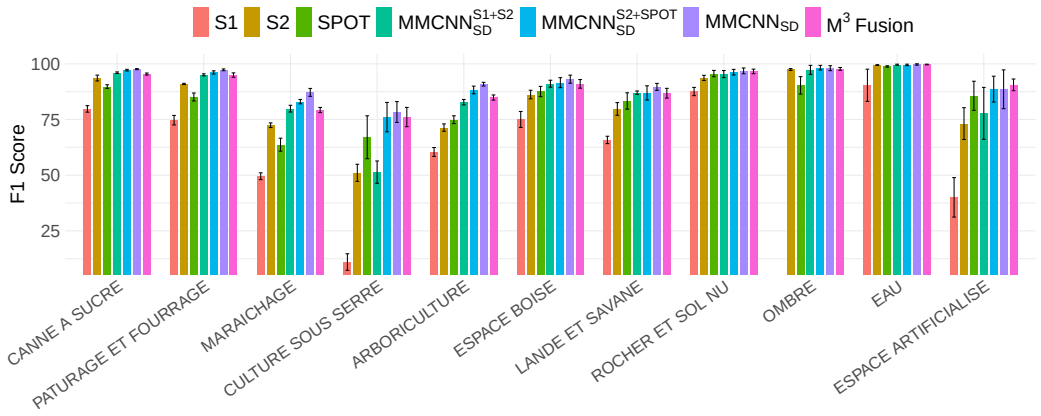


Figure 3.24 – Performances moyennes par classe d'occupation du sol sur l'île de la Réunion (l'écart-type est affichée sous forme de barre d'erreur)

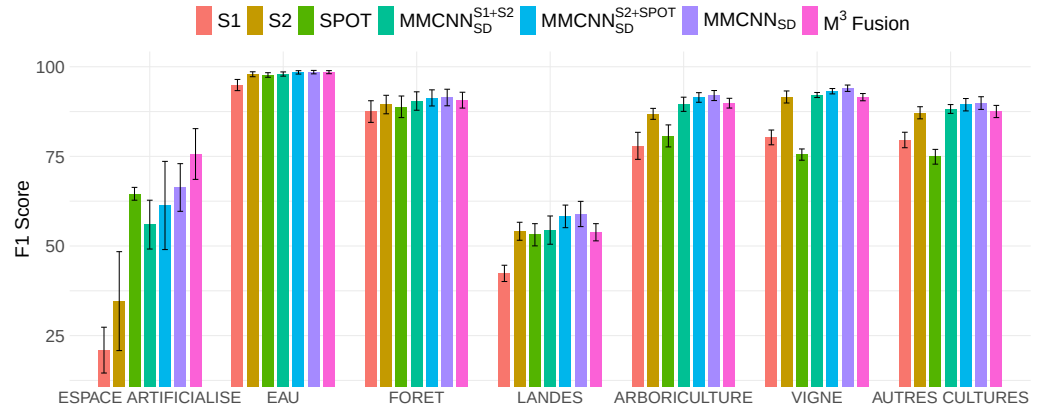


Figure 3.25 – Performances moyennes par classe d'occupation du sol sur le site de la Dordogne (l'écart-type est affichée sous forme de barre d'erreur)

aussi des patrons spatiaux, comme le montre l'amélioration des performances sur la classe des espaces artificialisés (zones urbaines entre autres).

3.3.4 Évaluation qualitative

Visualisation des représentations internes des CNNs

Dans cette étape de l'évaluation qualitative, nous avons visualisé les représentations internes apprises par les différents modèles sur les deux sites d'études. Pour ce faire, nous sélectionnons aléatoirement 300 échantillons par classe d'occupation du sol sur le jeu test et recueillons les représenta-

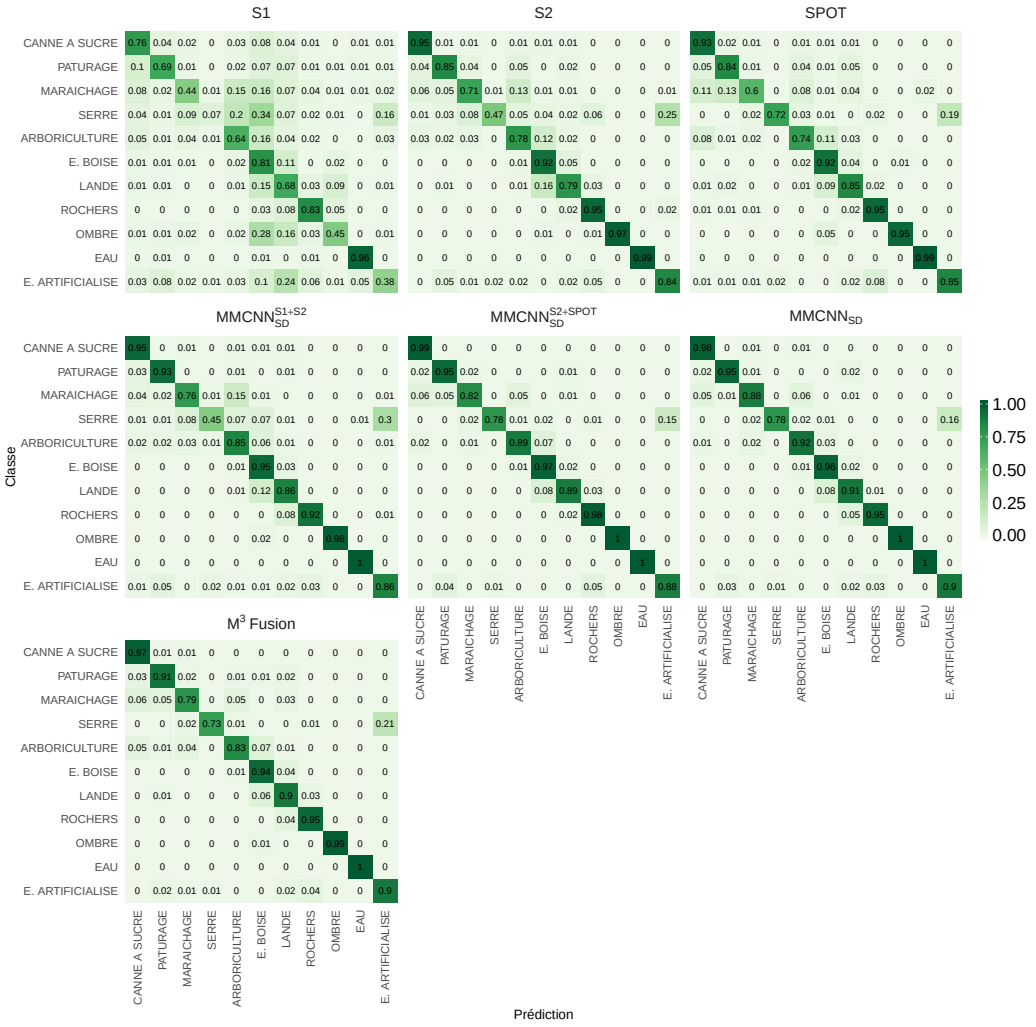


Figure 3.26 – Matrices de confusion entre les classes d'occupation du sol de l'île de la Réunion

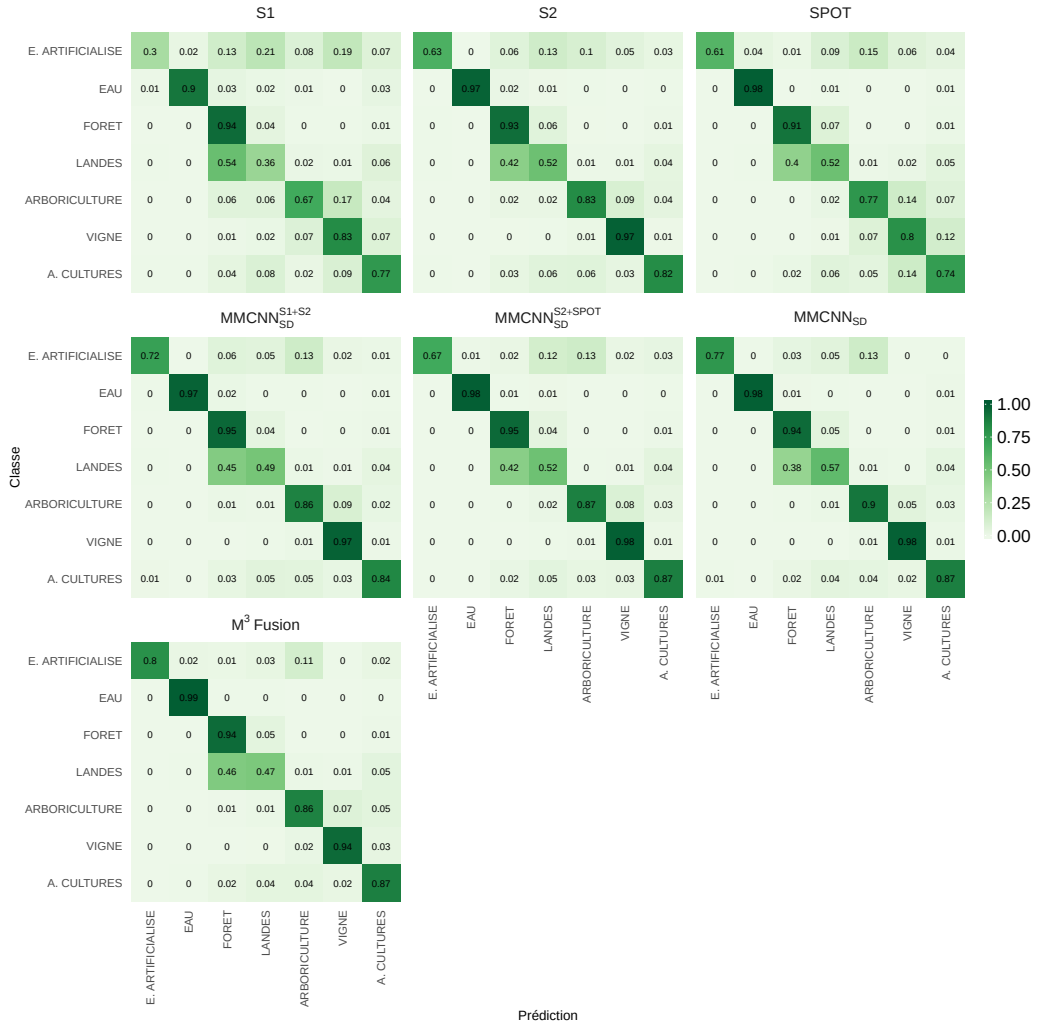


Figure 3.27 – Matrices de confusion entre les classes d'occupation du sol du site de la Dordogne

tions des différents modèles pour ces échantillons. Ensuite, nous appliquons l'algorithme t-SNE (t-distributed Stochastic Neighbor Embedding) (van der Maaten and Hinton, 2008) afin de réduire la dimensionnalité des caractéristiques en 2 axes à des fins de visualisation. Les résultats sont illustrées par les figures 3.28 et 3.29.

Sur les deux sites, nous pouvons observer une séparation des classes d'occupation du sol qui s'améliore à mesure que des sources additionnelles et complémentaires sont combinées. Comme observé dans l'évaluation quantitative, les données Sentinel-1 sont les moins discriminants d'entre les 3 sources

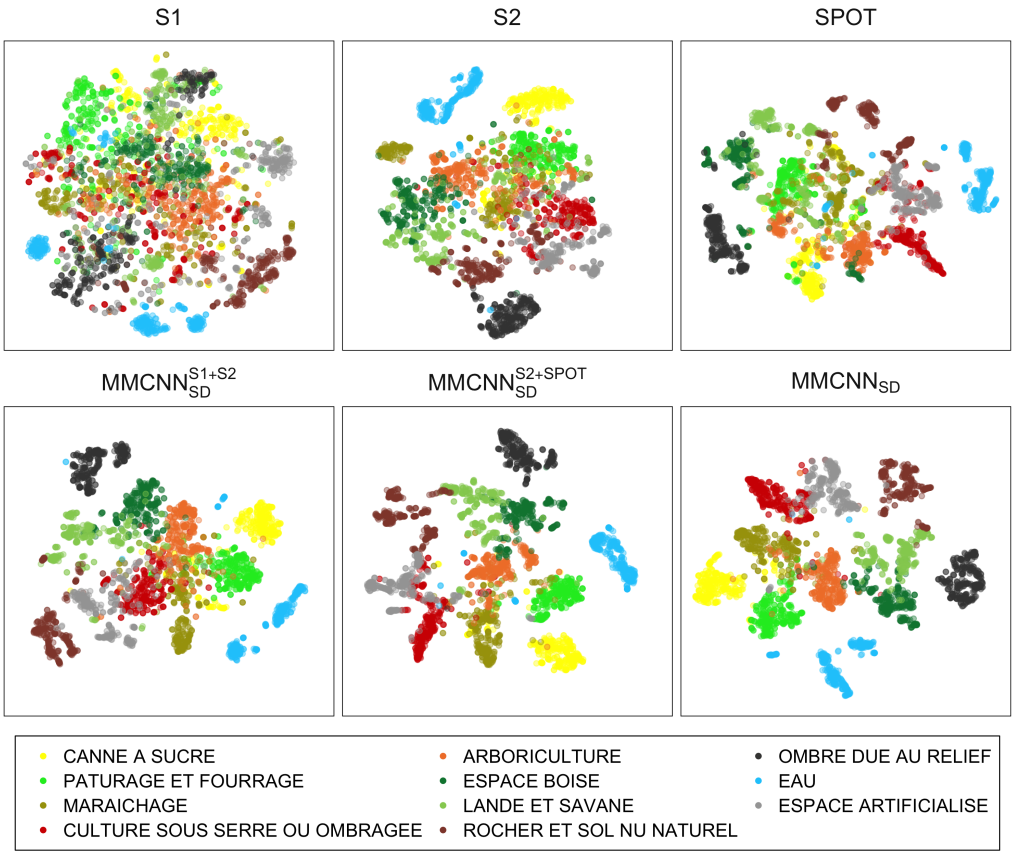


Figure 3.28 – Visualisation des représentations internes des modèles via t-SNE sur l'île de la Réunion

tandis que l'information spatiale à échelle fine fournie par SPOT est particulièrement utile sur l'île de la Réunion pour démêler les représentations des classes. Néanmoins, les représentations de certaines classes d'occupation sont encore difficilement séparables avec les données mono-modales en l'occurrence l'arboriculture et les espaces boisés ainsi que le maraîchage et les pâturage et fourrage sur l'île de la Réunion et les forêts et landes ou l'arboriculture, les vignes et les autres cultures sur la Dordogne. Ces ambiguïtés sont successivement allégées par la combinaison des données multi-modales particulièrement avec Sentinel-2 et SPOT et les trois sources associées, lesquelles permettent une séparation similaire des représentations tandis qu'avec Sentinel-1 et Sentinel-2 uniquement, certaines de ces confusions demeurent, notamment sur l'île de la Réunion. Dans l'ensemble, la visualisation des représentations internes apprises par les modèles s'accorde bien l'évaluation quantitative qui en a été faite précédemment.

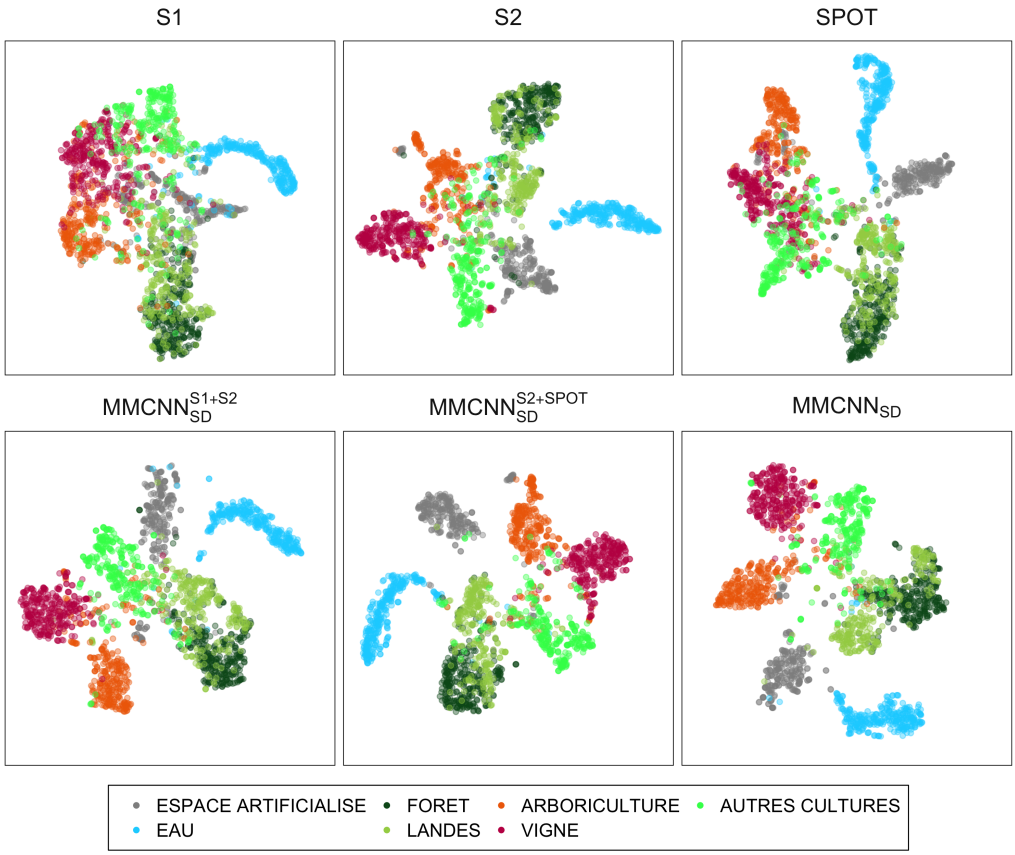


Figure 3.29 – Visualisation des représentations internes des modèles via t-SNE sur le site de la Dordogne

Extraits des cartes d'occupation du sol

En dernier lieu dans cette évaluation qualitative, nous nous intéressons aux cartes d'occupation du sol produites par les modèles sur l'un des sites en l'occurrence l'île de la Réunion. Ce dernier présente un paysage plus hétérogène et plus ardu en termes de classes d'occupation du sol que le site de la Dordogne. Rappelons que les cartes d'occupation du sol sont générées à la résolution des images Sentinel (10-m). Par ailleurs, étant donné que nous utilisons des imagettes pour décrire des emplacements géographiques spécifiques, les pixels en bordure (4 pixels dans chaque direction puisque la taille des imagettes Sentinel est 9×9) restent non étiquetés. Par souci de simplicité, nous ne présentons que les cartes produites à partir des combinaisons multimodales c'est-à-dire $MMCNN_{SD}^{S1+S2}$, $MMCNN_{SD}^{S2+SPOT}$ et $MMCNN_{SD}$. Nous présentons successivement cinq séries d'extraits des cartes dans la figure 3.30.

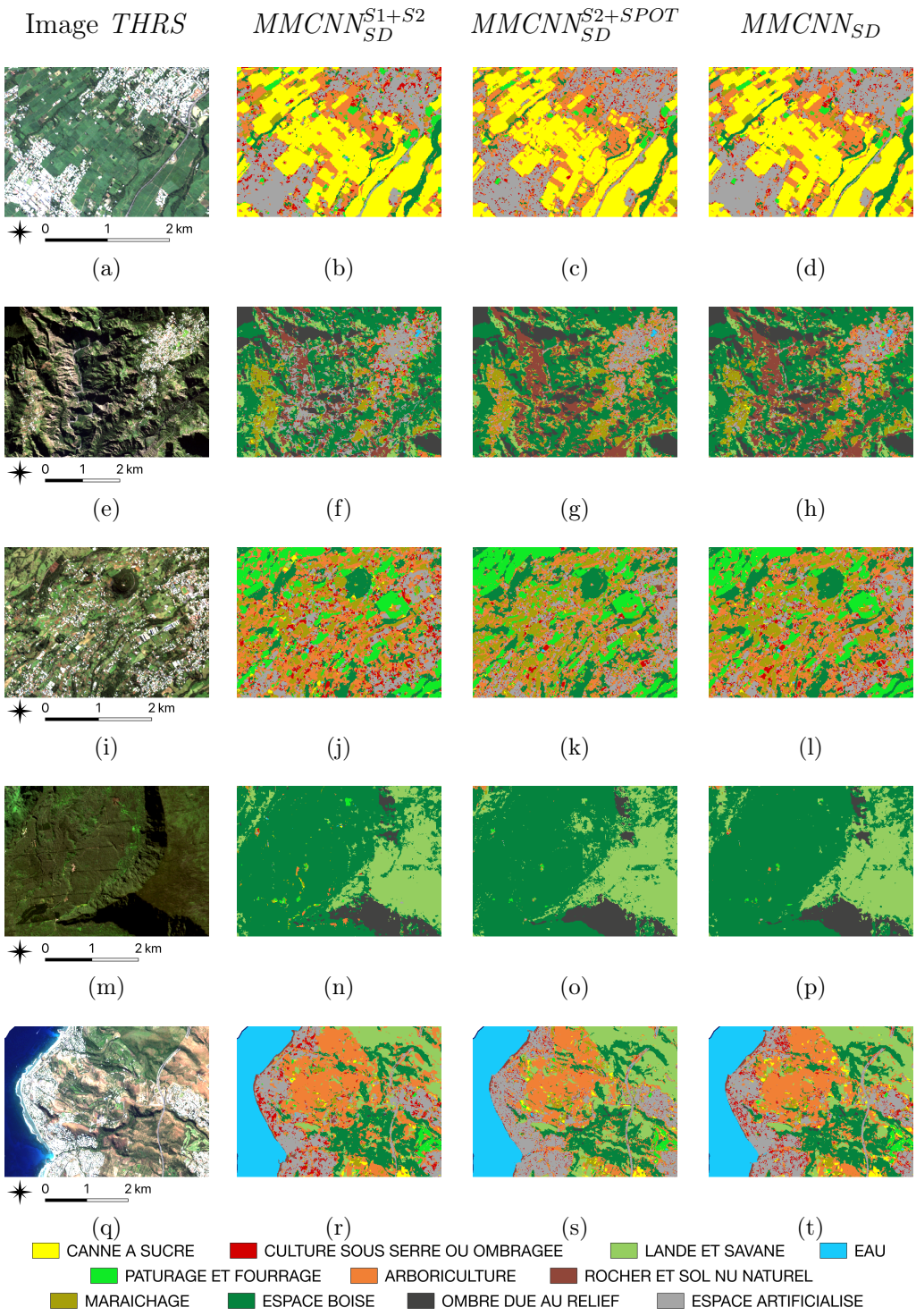


Figure 3.30 – Extraits des cartes d'occupation du sol sur l'île de la Réunion. L'image THRS est affichée en tant que référence.

La première série d'extraits (a-d) représente une partie de Saint-Pierre, un quartier urbain côtier avec des plantations de canne à sucre et d'arboriculture. Des erreurs de classification peuvent être mises en évidence entre espaces artificialisés et culture sous serre sur l'extrait de $MMCNN_{SD}^{S1+S2}$ tandis que l'introduction d'information spatiale à échelle fine (extraits de $MMCNN_{SD}^{S2+SPOT}$ et $MMCNN_{SD}$) a significativement réduit ces imprécisions. La seconde série d'extraits (e-h) représente une zone située dans le cirque de Cilaos, un paysage composé de hameaux entourés de quelques activités maraîchères. Ici la carte produite par $MMCNN_{SD}^{S1+S2}$ montre des erreurs de classification importantes entre la classe des rochers et sols nus naturels et les espaces artificialisés. Cet artefact est également légèrement perceptible sur l'extrait de $MMCNN_{SD}$ alors que la combinaison de Sentinel-2 et SPOT (c'est-à-dire $MMCNN_{SD}^{S2+SPOT}$) permet de mieux reconnaître les rochers et sols nus. La troisième série d'extrait (i-l) montre une zone autour du Tampon, un paysage mixte entre espaces urbains, pâturages et cultures maraîchères. Au-delà des confusions affichées par $MMCNN_{SD}^{S1+S2}$ entre les espaces artificialisés et les cultures sous serre, on note une surestimation générale des plantations de vergers bien que minimisée par $MMCNN_{SD}^{S2+SPOT}$ et $MMCNN_{SD}$. La quatrième série d'extraits (m-p) est localisée dans la forêt de Belouve. Le paysage est constitué de forêts primaires et de plantations forestières. Nous pouvons constater quelques légères imprécisions dans la détection de la forêt qui est mal classée avec l'arboriculture ainsi que les lande et savane. Ces imprécisions sont pour la plupart supprimées avec le modèle $MMCNN_{SD}$. Enfin, la cinquième série d'extraits est focalisée sur la région de Saint-Gilles les Bains. Le paysage est constitué d'arboriculture, de savanes, de quelques plantations de canne à sucre ainsi que de zones urbaines. Sur cet extrait, il y a une sous-estimation générale des landes et des savanes qui sont classées comme espaces boisées, bien que l'extrait de $MMCNN_{SD}^{S2+SPOT}$ l'atténue légèrement.

En somme, cette évaluation qualitative sous-tend également le bénéfice de combiner des données multi-modales de télédétection pour la cartographie de l'occupation du sol. Dans l'ensemble, les cartes d'occupation du sol produites par $MMCNN_{SD}^{S2+SPOT}$ et $MMCNN_{SD}$ sont d'une qualité satisfaisante, tandis que celle générée par $MMCNN_{SD}^{S1+S2}$ présente encore des erreurs extensives. Ceci est probablement dû au bruit subsistant dans les données SAR qui conduit parfois à des imprécisions telles que la surestimation de l'arboriculture et à la fine information spatiale fournie par l'image SPOT qui est particulièrement pertinente sur l'île de la Réunion.

3.4 Conclusion générale

Dans ce chapitre, nous avons traité de la caractérisation de l'occupation du sol en usant de méthodes d'apprentissage profond employant des données de télédétection multi-source. Deux méthodes sont proposées dans cette perspective.

La première méthode, dénommée *HOb2sRNN*, traite des séries temporelles radar et optique Sentinel, au niveau objet, à partir de RNNs équipés de mécanismes d'attention et incorpore dans le processus, des connaissances à priori sur les classes d'occupation du sol à travers une stratégie de pré-entraînement des paramètres. L'évaluation de la méthode *HOb2sRNN* a montré des résultats convaincants quant à la classification de l'occupation du sol dans des scénarios caractérisés par des quantités de données étiquetées variables. Sa comparaison avec plusieurs autres méthodes d'état de l'art met en évidence : (i) qu'elle est mieux adaptée qu'une approche standard du type forêts aléatoires communément employée pour la classification de l'occupation du sol à partir de séries temporelles ; (ii) que d'autres méthodes d'apprentissage automatique standards comme SVM ou MLP lui sont compétitives dans un scénario marqué par une quantité limitée de données étiquetées (site d'étude du Sénégal) et (iii) qu'elle est plus performante que l'adaptation dans un contexte multi-source du modèle TempCNN. Ce dernier résultat montre qu'au delà de la modélisation, par une approche RNN ou CNN, des dépendances qui caractérisent les séries temporelles radar et optique, d'autres caractéristiques méritent d'être prises en compte dans un cas multi-source par exemple la manière dont les données multi-sources sont associées ou encore les relations hiérarchiques entre les classes d'occupation du sol. À ce propos, les études d'ablation sur les diverses composantes du modèle *HOb2sRNN* montrent clairement leurs plus-values spécifiques dans l'amélioration des performances de classification. Par ailleurs, nous avions dans l'évaluation de la méthode *HOb2sRNN*, une contribution liée à l'IA explicable ou à l'interprétabilité des modèles d'apprentissage automatique souvent qualifiés de boites noires. Celle-ci a consisté en l'analyse qualitative des poids d'attention accordés aux différentes estampilles temporelles afin de mieux appréhender les décisions prises par le modèle. Nous avons ainsi pu établir des connexions entre ces décisions et des connaissances agronomiques dont nous disposions. Rappelons pour finir, que la méthode proposée est basée sur une approche orientée objet et de ce fait, est sujette à la qualité du processus de segmentation effectué en amont. Cette dernière peut constituer une source possible d'erreurs se propageant dans la cartographie de l'occupation du sol qui s'en suit. Néanmoins, il est hors du champ du travail réalisé dans cette thèse et centré sur le thème de l'occupation du sol et de l'ana-

lyse multi-source, d'étudier l'impact de cette source potentielle d'erreurs sur l'analyse ultérieure qui en découle. Ceci mériterait une autre étude dédiée et complète. Toutefois, nous pouvons affirmer avec conviction que les progrès réalisés dans l'analyse orientée objet des séries temporelles d'images satellitaires ne pourront être que bénéfiques à la méthode proposée (Ma et al., 2020).

La deuxième méthode, dénommée $MMCNN_{SD}$, explore quant-à-elle d'avantage la combinaison de données multi-sources et surtout multi-échelles de télédétection pour la caractérisation de l'occupation du sol. Elle associe pour ce faire aux séries temporelles radar et optique Sentinel précédentes à haute résolution spatiale, une image optique SPOT à très haute résolution spatiale. Cette méthode, qui traite les données au niveau pixel, repose sur une architecture à trois branches de CNNs, dont une dédiée à chaque source et pour laquelle une étude spécifique a permis de déterminer le type de réseau convolutif le plus adapté. Cette étude préalable a d'ailleurs montré que les convolutions 2D (spatiales) étaient largement plus efficaces que les convolutions 1D (temporelles) pour traiter les images radar, du fait qu'elles réduisent certainement plus le bruit spatial lié au speckle. De plus, la méthode $MMCNN_{SD}$ est dotée d'une stratégie d'auto-distillation permettant au modèle $MMCNN_{SD}$ de mieux combiner les caractéristiques par source en « apprenant de lui même ». Cette composante essentielle du modèle $MMCNN_{SD}$ améliore l'approche de combinaison multi-source équipant auparavant la méthode *HOb2sRNN* et qui repose sur des classifieurs auxiliaires par source. L'auto-distillation est utilisée dans ce cas précis pour permettre aux classifieurs auxiliaires par source de mimer la sortie du modèle et transférer ainsi les connaissances des couches les plus profondes vers les moins profondes. En somme, l'ajout d'information spatiale fine aux séries temporelles par le biais de la méthode $MMCNN_{SD}$ a montré également des résultats convaincants vis-à-vis de la caractérisation de l'occupation du sol, bien que le bénéfice soit variable en fonction des cas étudiés. Il est dès lors légitime de s'interroger sur la possibilité d'accroître les performances obtenues en associant d'autres sources de données pouvant être potentiellement un modèle numérique de surface pour prendre en compte les aspects de relief, les données radar en orbite descendante sans doute complémentaires de l'orbite ascendante mais non considérées dans cette thèse ou le reste des bandes spectrales optiques Sentinel-2 à 20-m de résolution spatiale dans d'autres domaines du spectre électromagnétique (red edge et infrarouge).

Chapitre 4

Suivi des rendements en petite agriculture familiale

Sommaire

4.1 Introduction

Le précédent chapitre a traité de méthodes d'apprentissage profond pour caractériser l'occupation du sol à partir de données de télédétection multi-source. Cette étape nous permet ainsi d'identifier les surfaces cultivées et de masquer les surfaces non cultivées en vue de la spatialisation de modèles de rendements agricoles qui font l'objet d'étude du présent chapitre. Plus précisément, dans ce chapitre, nous nous intéressons à l'estimation et à la prévision de rendements agricoles dans le cas de la petite agriculture familiale en Afrique subsaharienne.

Rappelons que dans les pays en développement d'Afrique subsaharienne, le secteur agricole demeure un important pilier de l'économie qui participe activement aux moyens de subsistance d'une grande part des populations notamment rurales. À ce jour, d'importants efforts en matière de politiques et d'investissements restent à faire afin de garantir un développement durable de l'agriculture et atteindre la sécurité alimentaire en Afrique subsaharienne. À cet effet, il est crucial de mettre en place un suivi précis des rendements des cultures vivrières qui sont encore très peu mesurés.

Les données de télédétection ont été employées dans diverses études ces dernières années avec le but de fournir des estimations quantitatives et/ou prévisions de rendements pour des cultures variées (Sakamoto et al., 2013; Kogan et al., 2013; Leroux et al., 2019). Jusque récemment, la plupart des initiatives reposaient sur des données optiques à basse résolution spatiale du type NOAA AVHRR (1-km) ou MODIS (250-m) (Rembold et al., 2013) et peu adaptée au contexte des parcelles de petite taille (moins d'un hectare) caractérisant les systèmes de culture de la petite agriculture familiale en Afrique subsaharienne. L'avènement de satellites optiques récents d'observation de la Terre aux résolutions spatiales améliorées comme Sentinel-2 (10-m) permet toutefois de se confronter plus efficacement à cet obstacle (Lambert et al., 2018). Néanmoins, très peu d'études (ex. Jin et al. (2019)) se sont intéressées au potentiel des données multi-source, particulièrement radar et optique, dans le contexte de la petite agriculture familiale en Afrique subsaharienne. L'association de ces sources est pourtant pertinente pour suivre efficacement le cycle de développement des cultures (Veloso et al., 2017), encore plus dans les climats tropicaux où la nébulosité fréquente compromet fortement l'acquisition d'images optiques claires à des étapes clés du développement des cultures. Ainsi, la disponibilité de données radar et optique quasi synchrones et à haute résolution spatiale comme les images Sentinel-1 et Sentinel-2 est une réelle opportunité pour la modélisation des rendements dans le cas de la petite agriculture familiale en Afrique subsaharienne.

Il est bien connu que des estimations et prévisions quantitatives de ren-

dements de cultures peuvent être obtenues en recourant d'une part à des modèles empiriques de régression ou d'autre part à des modèles de croissance de cultures (Rembold et al., 2013). Le premier groupe de méthodes cherche à établir une relation statistique entre des valeurs de rendements observés faisant office de références et des variables explicatives extraites souvent par télédétection telles que des indices de végétation (grâce à leur relation quasi-linéaire avec la biomasse (Tucker, 1979)) couplés parfois à des variables bioclimatiques (ex. température, précipitations) et des paramètres édaphiques (ex. teneur en carbone organique) (Leroux et al., 2020). Le second groupe de méthodes modélise de manière plus ou moins exhaustive la physiologie des cultures et simule leur croissance en recourant à des formules mathématiques et physiques relativement complexes. En raison de leur simplicité et de leur passage à l'échelle sur de vastes régions, les méthodes empiriques de régression ont été largement adoptées dans la littérature sur l'estimation et la prévision des rendements. Les approches courantes sont des modèles de régression linéaire univariés et multivariés (Jain et al., 2016; Jin et al., 2017; Lobell et al., 2019; Kim et al., 2019). Néanmoins, leur principale limite réside dans leur extrapolation à différentes années et à de nouvelles zones géographiques (Lobell, 2013). Pour pallier cet inconvénient, les techniques d'apprentissage automatique ont de plus en plus été considérées ces dernières années. C'est le cas des approches d'ensemble comme les forêts aléatoires (Kamir et al., 2020) ou les réseaux de neurones artificiels (Fieuzal et al., 2017; Fieuzal and Baup, 2017). Le principal avantage des techniques d'apprentissage automatique par rapport aux approches empiriques réside dans le fait qu'elles peuvent apprendre à prédire des valeurs cibles de rendements sans aucune des hypothèses préalables faites par les modèles paramétriques (Kamir et al., 2020). Dans cette thèse, nous traitons uniquement de méthodes empiriques et d'apprentissage automatique. De nos jours, l'utilisation des techniques d'apprentissage profond est également courante pour estimer et/ou prévoir les rendements (You et al., 2017; Khaki and Wang, 2019; Khaki et al., 2020; Wolanin et al., 2020; Khaki et al., 2021). Toutefois, ces dernières ont reçu très peu d'attention (Kaneko et al., 2017) pour la modélisation des rendements dans le cas de la petite agriculture familiale en Afrique subsaharienne, caractérisée certes par des données de référence (rendements observés) très limitées en raison des coûts de collecte importants, de l'inaccessibilité de certaines régions ou encore du manque de réglementation efficace sur des déclarations individuelles de production agricole.

Cette partie de la thèse traite donc de la modélisation des rendements dans le cas de la petite agriculture familiale en Afrique subsaharienne. Nous nous intéressons en particulier au mil qui est la principale culture vivrière de la région. Plus précisément, à partir de séries temporelles multi-sources (ra-

dar et optique) et dans un contexte très limité en données de référence, nous évaluons le potentiel des techniques d'apprentissage profond, comparés aux approches traditionnellement adoptées (méthodes empiriques et d'apprentissage automatique). Deux phases de modélisation sont conduites pour évaluer les rendements du mil : une phase d'estimation à la fin de le saison agricole et une phase de prévision en cours de saison. Les techniques de modélisation adoptées sont évaluées sur le site du bassin arachidier au Sénégal, pour lequel un total de 66 parcelles de mil ont été suivies pendant 3 saisons agricoles : 2017, 2018 et 2019 (voir la section 2.1.3 pour la description du jeu de données sur les rendements observés).

Ci-après, nous détaillerons les étapes de la méthodologie adoptée pour la modélisation des rendements dans le contexte décrit précédemment (section 4.2), puis nous présenterons et discuterons les résultats obtenus (section 4.3). Enfin, la section 4.4 conclura le chapitre.

4.2 Méthodologie

4.2.1 Variables explicatives

Afin de modéliser les rendements du mil sur notre site d'étude, nous recourrons à divers groupes de variables explicatives. Dans notre cas d'étude, les variables explicatives sont représentées par les valeurs observées dans les séries temporelles multi-sources Sentinel-1 et Sentinel-2. Les séries temporelles d'images sont collectées pour chacune des 3 saisons agricoles, sur la période de Juin à Décembre, avec un total de 17 images par source et par année. Nous avons collecté quelques images supplémentaires après la fin de la saison agricole (au plus tard en fin Octobre) afin d'obtenir un cycle complet de croissance des cultures (totalement ascendant et descendant). Les dates d'acquisition des images sont illustrées par la figure 4.1. Notons qu'à l'opposé des séries temporelles d'images Sentinel-1, les séries d'images Sentinel-2 sont irrégulièrement espacées au cours des trois saisons. Nous avons néanmoins conduit la phase d'estimation des rendements sans interpolation des données afin de réduire au maximum la complexité des modèles dans notre contexte d'étude limité en données d'apprentissage. Seule la phase de prévision des rendements, conduite avec des fenêtres temporelles de longueurs variées (voir ci-après section 4.2.3), est réalisée en interpolant les variables explicatives. À cette fin, les valeurs sont linéairement interpolées sur une grille régulière, chaque 5 jours, à partir de début Juin jusqu'à la fin Décembre.

Trois groupes de variables explicatives sont donc utilisées :

- les cœfficients de rétrodiffusion radar de surface en double polarisation

Figure 4.1 – Dates d'acquisition des images pluriannuelles Sentinel-1 et Sentinel-2. Les acquisitions Sentinel-2 sont éparses à cause de l'omniprésence de nuages.

(VV et VH)

- les bandes spectrales optiques du visible et de l'infrarouge à 10-m de résolution (au nombre de 4) et celles du red edge et de l'infrarouge moyen nativement à 20-m de résolution que nous avons rééchantillonnées à 10-m par la méthode du plus proche voisin (au nombre de 6)
- et enfin des indices de végétation au nombre de 7 dérivés des bandes optiques : NDVI, NDWI, MSAVI, EVI, GDVI, CIgreen et CIrededge (voir Tableau 1.4).

Les variables explicatives sont normalisées par année, dans l'intervalle [0,1], en considérant chaque série temporelle. La médiane des parcelles est ensuite extraite pour l'ensemble des variables afin de modéliser les rendements. Nous avons opté pour ce descripteur statistique car il sied au contexte de la petite agriculture en étant peu sensible aux valeurs extrêmes que peuvent présenter certains pixels au sein d'une même parcelle en raison de la variabilité intra-parcellaire induite par différentes pratiques culturales, la présence d'arbres ou encore des termitières (Félix et al., 2018).

4.2.2 Méthodes adoptées

Six méthodes sont comparées pour la modélisation des rendements du mil. Parmi elles, nous retrouvons quatre approches traditionnelles pour cette tâche (régression linéaire, régression ridge, fôrets aléatoires et perceptron multi-couche) et deux méthodes d'apprentissage profond (**LSTM** et **CNN**).

La régression linéaire (**LR**) fait figure d'approche de base ou référence. Nous l'avons employé en utilisant comme unique variable explicative la valeur maximale de NDVI au cours de la saison agricole. Le potentiel de cette approche a déjà été démontré par différentes études dans des contextes agricoles similaires (Azzari et al., 2017; Jain et al., 2016; Lambert et al., 2018). Étant donné que la majorité des rendements des cultures céréalières est déterminée par l'activité photosynthétique pour une période allant de la floraison jusqu'à la récolte (Rembold et al., 2013), la valeur de NDVI au pic de la saison est considérée comme une variable explicative des rendements finaux.

La Régression ridge (**RR**) (Hoerl and Kennard, 1970) est technique de régression multiple et à la différence de la régression linéaire, nous l'employons avec les variables explicatives multi-sources en faisant varier le terme de pénalité (norme L^2) mis pour régulariser les cœfficients de régression afin de prévenir la multicolinéarité et réduire la complexité du modèle.

Les Forêts aléatoires (**RF**) (Breiman, 2001) représentent l'une des méthodes les plus communes – pour ne pas dire la méthode la plus fréquente – pour l'estimation et la prévision des rendements. Nous l'employons également avec les variables explicatives multi-sources en faisant varier certains hyperparamètres importants que sont le nombre d'arbres de décision, la profondeur maximale des arbres ou encore le nombre maximum de caractéristiques ou variables.

Le Perceptron multi-couche (**MLP**) (Rumelhart et al., 1986) est aussi parfois désigné sous l'appellation de réseaux de neurones artificiels dans la littérature sur l'estimation et la prévision des rendements. Dans notre contexte où les données d'entraînement sont très limitées, nous avons mis au point un modèle MLP très léger. Il est constitué de deux couches cachées avec 64 neurones chacune. Chaque couche est associée à une fonction d'activation ReLU.

Les techniques d'apprentissage profond que nous comparons à ces approches traditionnelles sont donc des modèles LSTM et CNN. Tout comme pour le MLP, nous avons adopté des architectures très légères. Ainsi, le modèle LSTM est constitué de 64 unités cachées. Pour concevoir l'architecture du réseau CNN, nous avons suivi certaines directives générales adoptées dans la littérature où le nombre de filtres augmente au fur et à mesure que le réseau devient profond. L'architecture du CNN se compose de 6 couches de convolutions 1D (temporelles) avec au maximum 64 filtres suivies d'une couche de Pooling. Le reste des détails architecturaux est consigné dans le tableau 4.1. Nous employons la fonction de coût Huber [1] et l'optimiseur Adam (Kingma and Ba, 2015) pour apprendre les paramètres des réseaux de neurones. Étant

1. https://en.wikipedia.org/wiki/Huber_loss

donné l'aspect critique que revêt l'apprentissage de ces paramètres en raison du peu d'échantillons d'entraînement à disposition, nous employons la technique de la moyenne mobile afin d'obtenir les poids finaux des réseaux.

Nous reportons dans le tableau 4.2, les hyper-paramètres et valeurs associées des méthodes évaluées. Notons que les modèles RR et RF sont optimisés en variant leur hyper-paramètres dans une gamme de valeurs associées tandis que ceux des réseaux de neurones sont empiriquement fixés. Les approches LR, RR et RF sont implémentées en utilisant la bibliothèque Python Scikit-learn tandis que les réseaux de neurones (MLP, LSTM et CNN) sont implémentés avec la bibliothèque Python Tensorflow.

Tableau 4.1 – Détails sur l'architecture du réseau CNN. Conv désigne une couche de convolution, nf est le nombre de filtre, k la taille du filtre, s est le pas et act est la fonction d'activation. Comme pour les autres réseaux de neurones, nous utilisons au maximum 64 filtres de convolution.

Couche 1	Conv(nf=16, k=5x1, s=1, act=ReLU)
Couche 2	Conv(nf=16, k=3x1, s=1, act=ReLU)
Couche 3	Conv(nf=32, k=3x1, s=2, act=ReLU)
Couche 4	Conv(nf=32, k=3x1, s=1, act=ReLU)
Couche 5	Conv(nf=64, k=1x1, s=1, act=ReLU)
Couche 6	Conv(nf=64, k=1x1, s=1, act=ReLU)
Couche 7	GlobMaxPooling1D

Tableau 4.2 – Hyper-paramètres et valeurs associées des méthodes évaluées.

Modèle	Hyper-paramètres	Valeur ou Gamme
RR	Pénalité	{0.001,0.1,1}
RF	Nombre d'arbres	{50, 100, 200}
	Profondeur maximale	{10, 20, 50, None}
	Maximum de caractéristiques	{'sqrt', 'log2', None}
MLP LSTM CNN	Taux d'apprentissage	1×10^{-3}
	Taille par lot	1
	Nombre d'époques	1000

4.2.3 Phases de modélisation

Nous avons considéré deux phases de modélisation des rendements : une phase d'estimation et une phase de prévision. La phase d'estimation, hormis

pour l'approche de base (LR), est basée sur les données multi-sources d'origine c'est-à-dire les séries temporelles non interpolées avec 17 acquisitions par source et par année. Les rendements de mil sont tout d'abord estimés en utilisant individuellement les différents groupes de variables prédictives. Ensuite, nous évaluons des scénarios de combinaison multi-source où les groupes de variables explicatives sont concaténés. Nous avons analysé quatre scénarios multi-sources possibles : (i) Radar (SAR) et Bandes optiques (Opt.), (ii) SAR et Indices de végétation (VI), (iii) Opt. et VI et (iv) l'ensemble des 3 groupes (SAR, Opt. et VI). Finalement, le meilleur modèle de la phase d'estimation est exploré ultérieurement dans la phase de prévision.

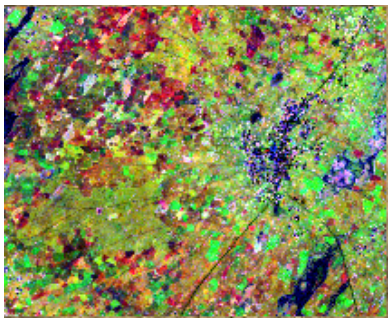
La phase de prévision des rendements est basée sur les données interpolées avec un pas de 5 jours. Elle est conduite sur une période s'étalant du début des acquisitions en début Juin à la fin de la saison agricole au plus tard en fin Octobre. Différentes fenêtres temporelles sont alors considérées afin de déterminer une période adéquate avant la récolte où des prévisions de rendement satisfaisantes pourraient être obtenues. Les rendements de mil sont ainsi prédits dès la phase d'émergence en début Juillet à la phase de sénescence en considérant des fenêtres temporelles raccourcies par incréments de 15 jours à partir du début. En général, l'intervalle de 15 jours est considéré comme une période de temps raisonnable pour capturer des changements significatifs dans le développement de la végétation.

4.2.4 Évaluation des modèles

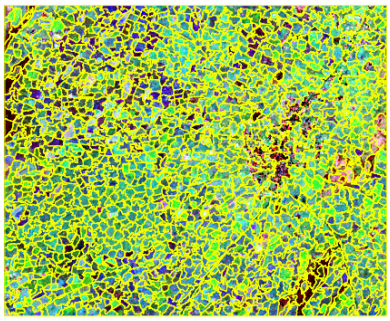
Les performances des modèles sont évaluées à travers une procédure de validation-croisée à 3 blocs répétée aléatoirement 10 fois. Un quart des blocs d'entraînement (environ 11 échantillons sur 44 parcelles) est utilisé pour valider les modèles évaluées c'est-à-dire sélectionner les valeurs associées aux hyper-paramètres ou sauvegarder les paramètres ou poids des réseaux de neurones. Les prédictions sont alors faites sur les blocs de test (environ 22 parcelles) et finalement les performances des modèles sont moyennées et reportés en considérant les métriques R^2, RMSE et MAE (voir section 1.2.3 pour les définitions.)

4.2.5 Spatialisation des rendements du mil

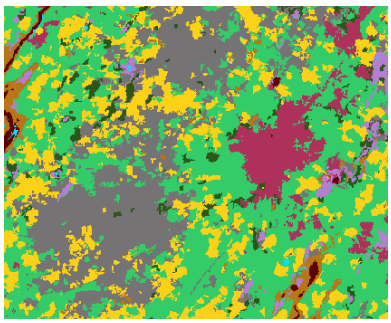
Dans le but de spatialiser les rendements du mil à l'échelle de la parcelle sur la zone d'étude, nous avons généré des pseudo parcelles sur l'étendue du site, en effectuant d'abord une segmentation puis en masquant les superficies non cultivées des objets grâce à l'occupation du sol. Pour ce faire, nous avons produit successivement les cartes d'occupation du sol du site d'étude pour



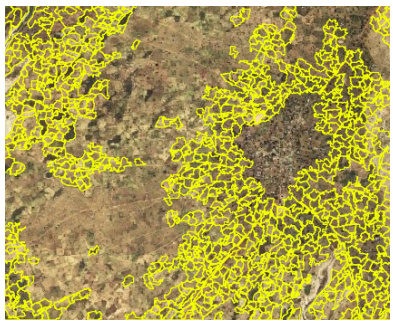
(a) Série temporelle de NDVI



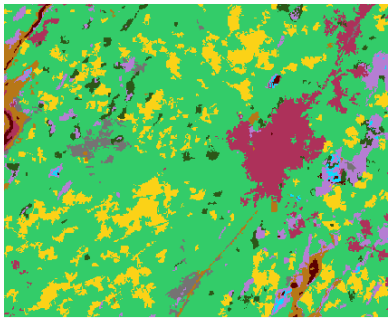
(b) Pseudo parcelles candidates



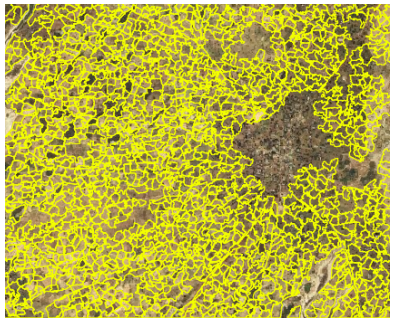
(c) Occupation du sol 2018



(d) Pseudo parcelles 2018



(e) Occupation du sol 2019



(f) Pseudo parcelles 2019

(g) Classes d'occupation du sol

Figure 4.2 – Processus de génération des pseudo parcelles pour la spatialisation des rendements du mil en 2018 et 2019 (le fond de carte utilisé pour les sous-figures 4.2d et 4.2f est issu de Bing Maps).

l'année 2018 et 2019 en employant la méthode *HOb2sRNN* (voir section 3.2). Celle de 2018, qui a fait l'objet d'une évaluation dans la section 3.2, a été produite avec une précision globale de 90% et un F1-score de 88% sur la classe céréale (majoritairement du mil). La carte de 2019 a été produite de manière analogue en utilisant la vérité terrain de 2019 (voir tableau 2.3) avec une précision globale de 79% et un F1-score de 77% sur la classe céréale. En raison de contraintes budgétaires, aucune donnée sur l'occupation du sol en 2017 n'a pu être collectée. Ainsi, nous nous focalisons uniquement sur la spatialisation des rendements pour les saisons agricoles 2018 et 2019.

La segmentation de la zone d'étude est effectuée via l'algorithme SLIC (Achanta et al., 2012) à partir d'une série temporelle de NDVI extraite d'images PlanetScope de la saison 2018 (Figures 4.2a et 4.2b). Cette étape nous permet d'obtenir des pseudo parcelles candidates de taille comparables ayant des caractéristiques biophysiques semblables. Ensuite pour chaque saison culturale entre 2018 et 2019, une intersection spatiale est réalisée entre les pseudo parcelles candidates et la carte d'occupation du sol correspondante (Figures 4.2c et 4.2e) produite au niveau objet. De cette façon une classe d'occupation du sol est attribuée à chaque résultat de l'intersection. Finalement, nous obtenons les pseudo parcelles spécifiques à la saison agricole par vote majoritaire sur la classe d'occupation du sol des parcelles candidates, masquant ainsi les superficies non cultivées et autres que des céréales (Figures 4.2d et 4.2f).

4.3 Résultats et Discussion

4.3.1 Résultats

Phase d'estimation des rendements

Nous présentons tout d'abord les résultats de la phase d'estimation des rendements avec les groupes distincts de variables explicatives. Les performances moyennes des modèles, suivant la procédure de validation croisée, sont reportées dans le tableau 4.3. Nous notons tout d'abord les mauvaises performances, par rapport au reste des méthodes, de l'approche de base employant une régression linéaire avec la valeur maximale de NDVI observée au cours du développement des cultures comme variable explicative (R^2 moyen de 0.08). Indépendamment des variables explicatives, les modèles RF et MLP ont obtenu des performances comparables et meilleures que celles des autres approches (R^2 moyen de 0.48) en l'occurrence celles d'apprentissage profond (R^2 moyen de 0.42) et la régression ridge (R^2 moyen de 0.41). À propos des différents groupes de variables explicatives, nous notons que les variables ra-

dar sont moins efficaces que les autres variables optiques (bandes optiques ou indices de végétation) sur notre site d'étude, ce qui peut être particulièrement mis en lumière pour le modèle CNN (R^2 moyen de 0.06). De plus, le comportement des modèles est variable vis-à-vis des bandes optiques ou indices de végétation. Certains modèles (RR, MLP) sont plus performants avec les bandes optiques tandis que c'est l'inverse pour les autres (RF, LSTM, CNN) qui le sont plutôt avec les indices de végétation. Dans l'ensemble, ce sont les modèles RF et MLP entraînés respectivement avec les indices de végétation et les bandes optiques qui présentent les meilleurs résultats de cette phase d'estimation obtenant respectivement des RMSE moyens similaires de 446 kg/ha et 445 kg/ha.

Tableau 4.3 – Performances moyennes des modèles

Max. NDVI	R^2	MAE	RMSE
LR	0.08 ± 0.13	473.27 ± 44.32	593.66 ± 63.31
SAR	R^2	MAE	RMSE
RR	0.24 ± 0.14	422.28 ± 34.94	539.98 ± 64.08
RF	**0.32** ± 0.15	400.39 ± 43.70	**509.53** ± 71.23
MLP	0.20 ± 0.26	409.49 ± 54.39	547.38 ± 100.18
LSTM	0.25 ± 0.19	**397.37** ± 46.47	533.81 ± 85.29
CNN	0.06 ± 0.23	451.25 ± 55.23	597.27 ± 86.67
Optique	R^2	MAE	RMSE
RR	0.41 ± 0.20	360.65 ± 55.26	471.03 ± 90.99
RF	0.44 ± 0.12	361.5 ± 37.14	460.83 ± 68.65
MLP	**0.48** ± 0.17	**335.5** ± 52.24	**445.0**± 87.85
LSTM	0.33 ± 0.19	391.58 ± 51.42	500.68 ± 71.15
CNN	0.35 ± 0.17	377.28 ± 60.70	497.10 ± 84.35
VI	R^2	MAE	RMSE
RR	0.36 ± 0.17	381.29 ± 55.52	492.39 ± 86.52
RF	**0.48** ± 0.13	**346.06** ± 41.0	**446.32** ± 67.77
MLP	0.42 ± 0.14	348.41 ± 40.68	471.15 ± 73.12
LSTM	0.42 ± 0.12	358.60 ± 48.88	468.92 ± 68.97
CNN	0.42 ± 0.21	360.11 ± 67.56	468.91 ± 94.68

Nous illustrons sur la figure 4.3, les diagrammes de dispersion représentant les valeurs observées et prédites des rendements, moyennées à travers de la procédure de validation croisée pour l'ensemble des modèles. En considérant les prédictions moyennées, nous remarquons que le modèle CNN entraîné avec les indices de végétation, noté CNN (VI), montre de meilleurs accords que les autres approches entre observations et prédictions (R^2=0.58), par exemple RF (VI) avec un R^2 de 0.53 ou MLP (Opt.) avec un R^2 de 0.54.

Toutefois, les diagrammes en boîte marginaux montrent pour l'ensemble des modèles, une sous-estimation des valeurs élevées ainsi qu'une surestimation des valeurs faibles de rendements.

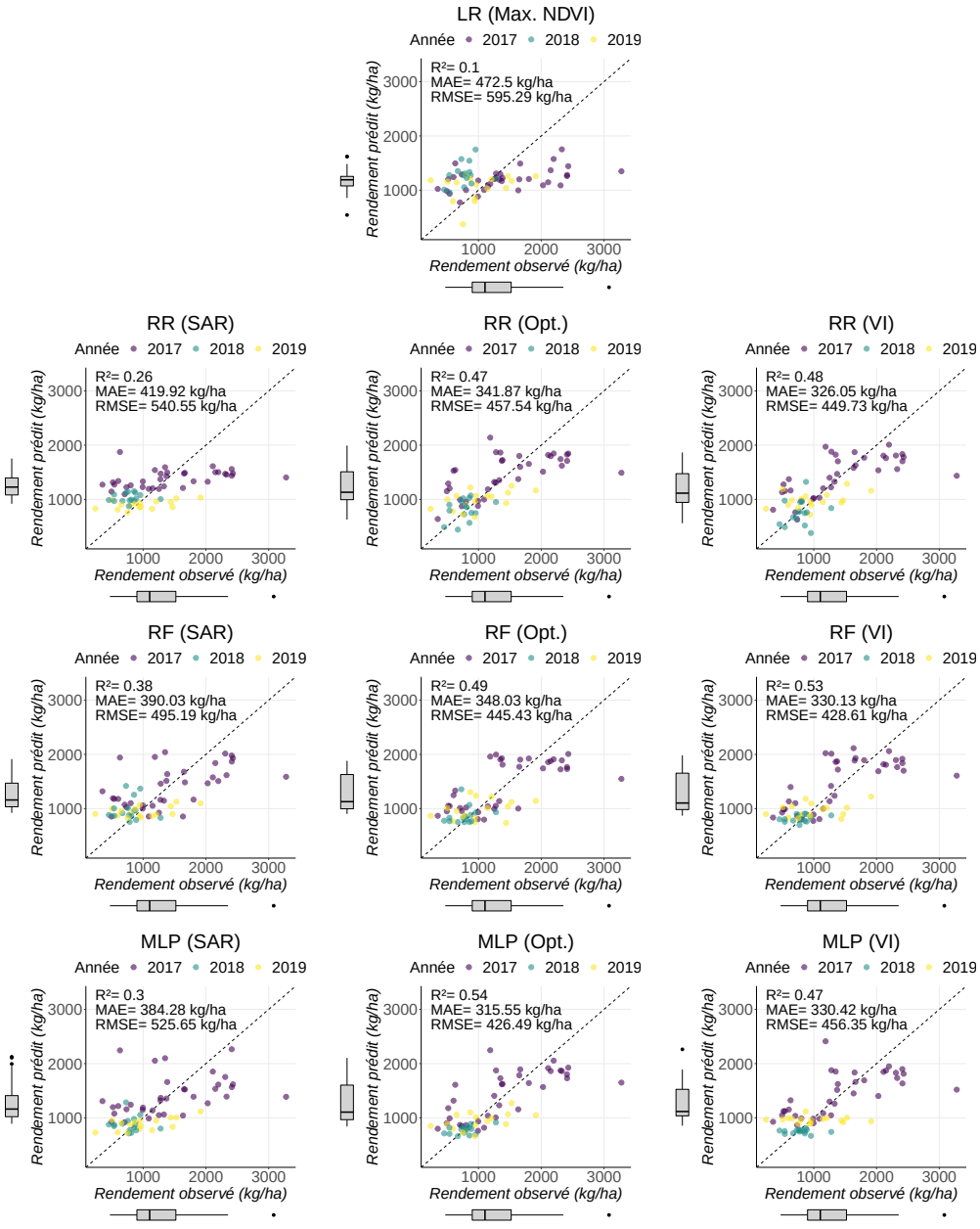


Figure 4.3 – *Continue*

Pour résumer cette phase d'estimation des rendements à partir des groupes distincts de variables explicatives, nous avons observé que les approches tradi-

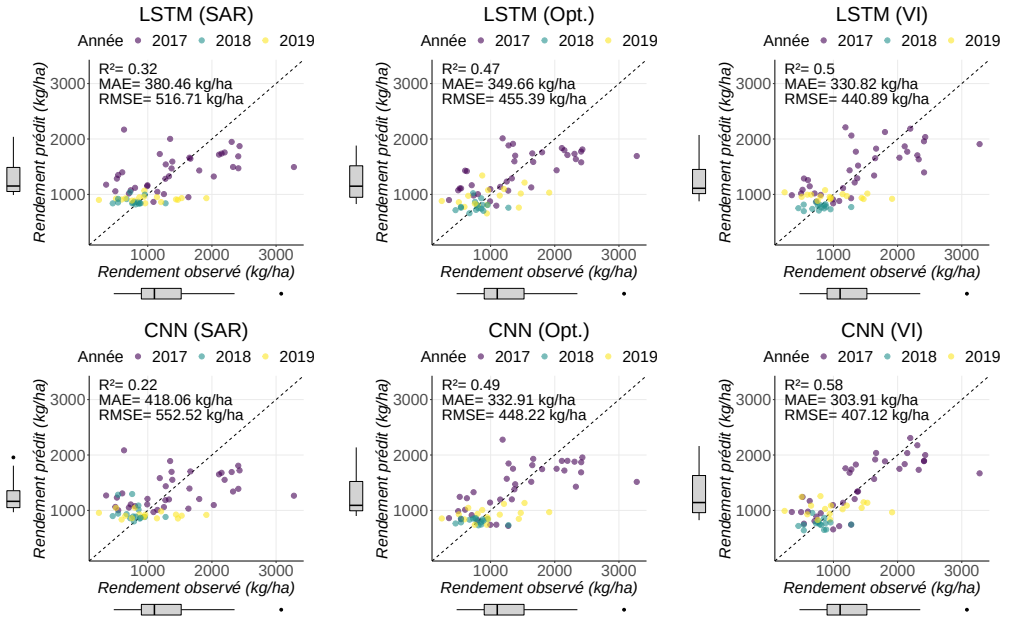


Figure 4.3 – Diagrammes de dispersion représentant les observations et prédictions de rendements moyennées à travers la procédure de validation croisée.

tionnelles telles que RF ou MLP restent, dans ce contexte très limité en données d'apprentissage, plus performantes que les techniques d'apprentissage profond, même si le comportement moyen du modèle CNN est prometteur.

Par la suite, nous avons analysé les performances des modèles pour l'estimation des rendements en considérant les quatre scénarios de combinaison des variables explicatives présentés en section 4.2.3 : (i) SAR et Opt., (ii) SAR et VI, (iii) Opt. et VI et (iv) l'ensemble des 3 groupes (SAR, Opt. et VI). Les performances obtenues avec la combinaison multi-source sont illustrées sur la figure 4.4. À titre de comparaison, nous y avons également reporté les meilleures performances précédentes (MPP), obtenues pour chaque modèle en utilisant les groupes distincts (soit les bandes optiques ou les indices de végétation). À l'exception des approches RF et CNN, les performances des modèles restent similaires ou sont plus faibles par rapport à celles obtenues précédemment. Nous notons respectivement des améliorations du R^2 de 0.48 à 0.51 pour le modèle RF et 0.42 à 0.46 pour le CNN. Cette amélioration pour les deux modèles est relative au scénario (ii) SAR et VI. Ce scénario multi-source est également le plus efficace d'entre les quatre pour le modèle LSTM bien qu'il reste moins performant que le scénario mono-source. Dans l'ensemble, la combinaison multi-source (radar et optique) n'améliore pas

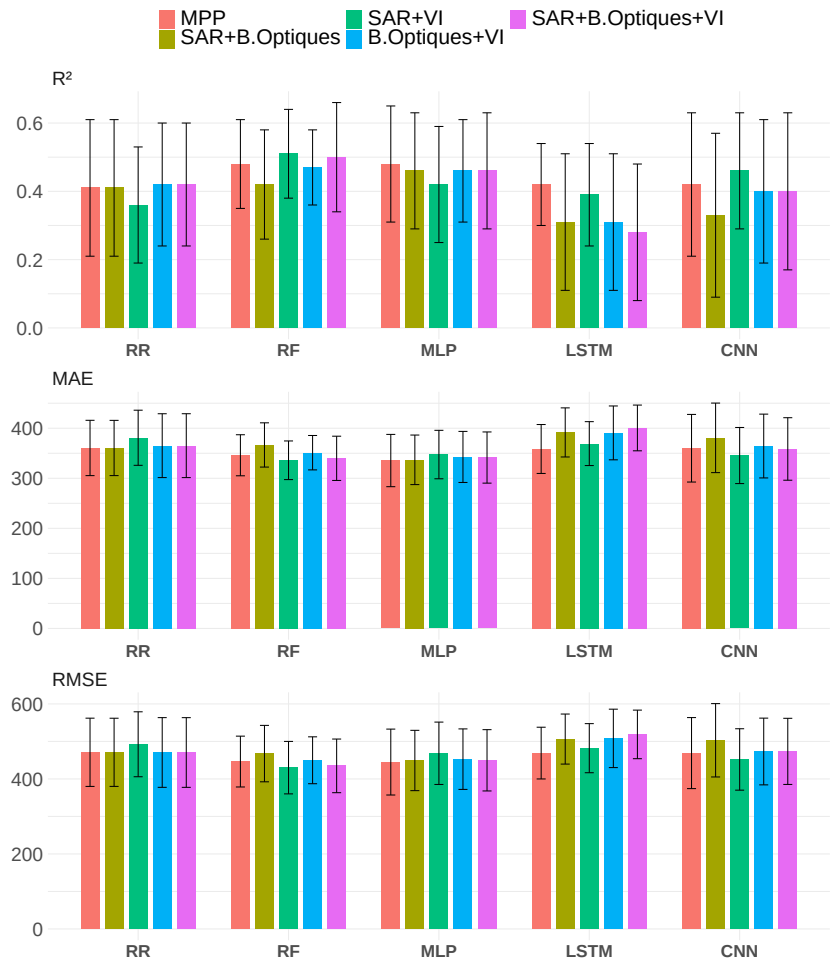


Figure 4.4 – Performances moyennes et écart type (sous forme de barre d'erreur) obtenues avec la combinaison des différents groupes de variables explicatives. Nous notons MPP les meilleures performances précédentes des modèles obtenues avec les groupes distincts de variables explicatives.

systématiquement l'estimation des rendements du mil sur notre site d'étude. Les améliorations que nous avons constaté semblent toutefois étroitement liées à l'approche de modélisation utilisée.

Phase de prévision des rendements

Dans cette phase de prévision de rendements, nous employons le modèle RF et la combinaison multi-source radar et indices de végétation qui semble plus adapté que les autres scénarios testés pour la modélisation des rendements du mil sur notre site. Les résultats sont illustrés par la figure 4.5. Pour rappel, la prévision des rendements est effectuée à partir du stade d'émergence des cultures jusqu'à leur sénescence, en considérant différentes fenêtres temporelles raccourcies par incréments de 15 jours.

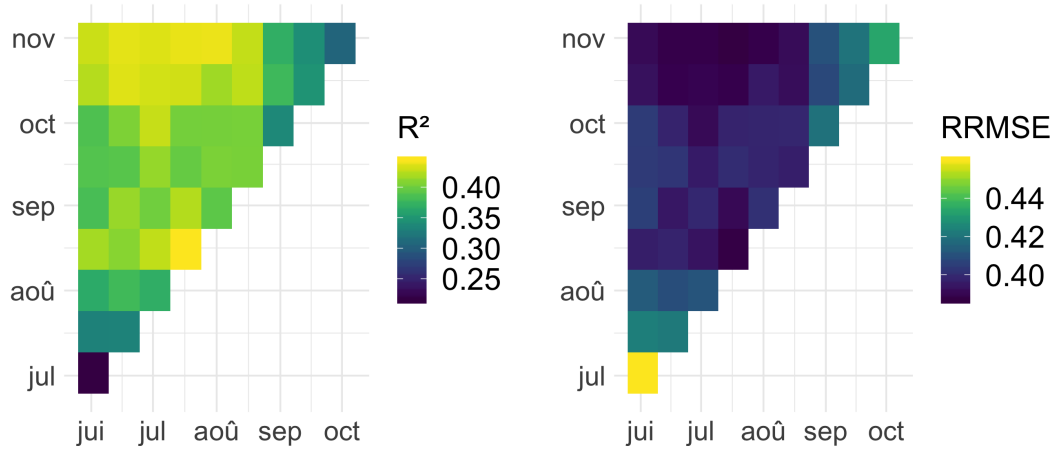


Figure 4.5 – Performances moyennes obtenues en considérant le R^2 et le RRMSE (ou RMSE relatif, soit la valeur du RMSE divisée par la moyenne des observations) durant la phase de prévision des rendements. L'axe x représente le début des fenêtres temporelles de prévision et l'axe y sa fin.

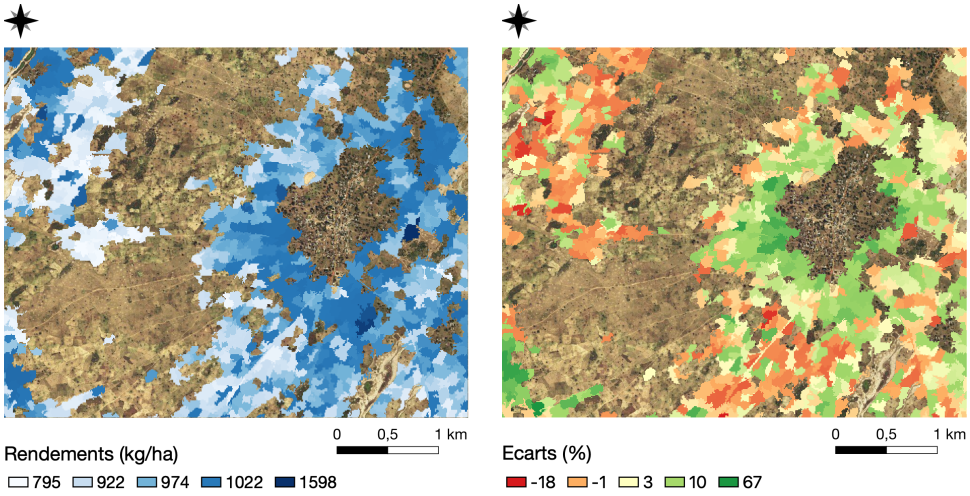
Nous remarquons que les prévisions deviennent plus précises à mesure que les fenêtres temporelles considérées s'approchent de la période de récolte vers fin Octobre au plus tard. Des performances de prévision plus ou moins similaires à celles de l'estimation sont obtenues pour une période s'étendant entre la mi-Août et la fin Octobre. Bien que la fenêtre temporelle s'étendant entre la mi-Juillet et la mi-Août affiche les meilleures performances en prévision (R^2 de 0.44), des prédictions stables sont obtenues à partir de la mi-Octobre (R^2 de 0.43), soit deux semaines avant la fin de toute activité agricole. Par ailleurs, la longueur des fenêtres temporelles considérées, ne semble

pas influencer les prédictions à la mi-Octobre, puisque les performances obtenues sont sensiblement les mêmes pour les périodes considérées. Notons tout de même la chute en précision des prévisions considérant des fenêtres temporelles tardives (Septembre/Octobre – Octobre/Novembre) qui s'explique certainement par le fait que leurs prédictions soient basées uniquement sur des informations de sénescence des plantes.

Spatialisation des rendements

Dans cette dernière partie, nous présentons les résultats de la spatialisation des rendements pour les pseudo parcelles de mil générées en 2018 et 2019. Nous employons à nouveau le modèle RF et la combinaison multi-source SAR et VI. La spatialisation est conduite pour les deux phases de modélisation (estimation et prévision). De ce fait, nous évaluons aussi la distribution spatiale des écarts entre prévisions de rendements en cours de saison (à peu près deux semaines avant la fin des récoltes) et estimations en fin de saison ainsi que leurs tendances (surestimations ou sous-estimations). Les différentes cartes sont présentées à la figure 4.6.

Sur les cartes de rendements des deux années (Figures 4.6a et 4.6c), nous notons qu'il existe une variabilité spatiale notable, même pour des pseudo parcelles contiguës. Les moyennes des rendements estimés sont respectivement de 965 kg/ha et 1080 kg/ha en 2018 et 2019, avec des cœfficients de variation de 8% et 17%. Sans grande surprise, les rendements les plus élevés sont obtenus dans l'anneau de fertilité, qui désigne les parcelles entourant des zones habitées (villages). Ce sont les parcelles qui reçoivent le plus souvent des fertilisants (ex. fumiers). Ainsi, les parcelles plus éloignées et peu fertilisées obtiennent souvent des rendements plus faibles. C'est l'exemple des parcelles en jachère en 2018 (toute la partie à l'ouest) et cultivées en 2019 dont les rendements sont parmi les plus faibles. En ce qui concerne la distribution spatiale des écarts entre prévisions et estimations sur les deux années (Figures 4.6b et 4.6d) et plus précisément leur tendances, nous notons clairement une surestimation d'ensemble des rendements en 2018, notamment dans l'anneau de fertilité, tandis qu'en 2019 les rendements sont plutôt sous-estimés pour cette zone particulière. Les pourcentages moyens des écarts sont d'environ 5% en 2018 et de -0,5% en 2019. De manière globale, il n'y a pas de tendance claire d'une année à l'autre en ce qui concerne les écarts obtenus pour une même zone. De plus la plupart des écarts restent faibles comme le montrent les courbes de densité pour les deux années (figure 4.7), ce qui indique des capacités de prévision satisfaisantes par rapport aux estimations.



(a) Estimations des rendements 2018

(b) Écarts (prévisions-estimations) 2018

(c) Estimations des rendements 2019

(d) Écarts (prévisions-estimations) 2019

Figure 4.6 – Cartographie des rendements du mil et écarts entre prévisions et estimations pour les saisons culturales 2018 et 2019 (une discrétisation par quantiles est appliquée pour les différentes cartes ; le fond de carte utilisé pour les sous-figures est issu de Bing Maps).

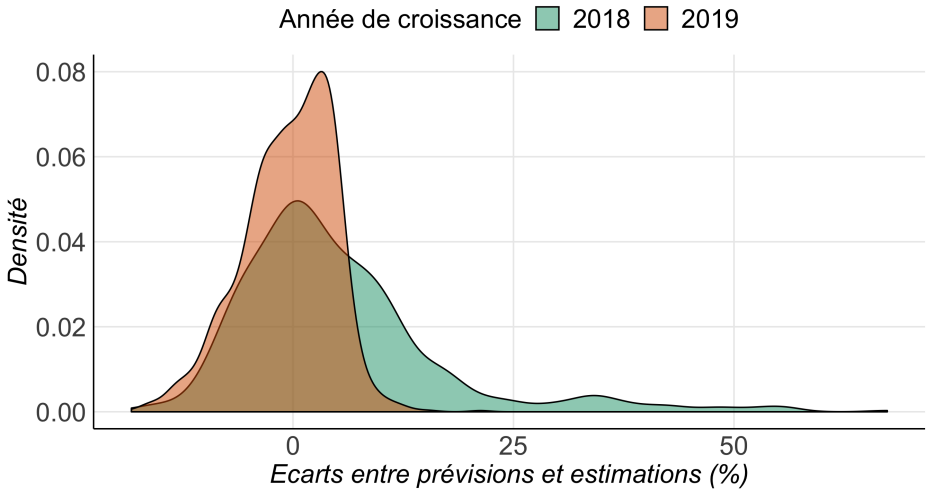


Figure 4.7 – Courbes de densité représentant les écarts entre prévisions et estimations pour les saisons culturales 2018 et 2019.

4.3.2 Discussion

Dans ce travail, nous avons comparé diverses approches pour la modélisation des rendements du mil. Parmi elles, nous avions comme approche de base ou référence un modèle de régression linéaire avec comme unique variable explicative la valeur maximale de NDVI observée au cours du développement des cultures. Bien que le potentiel de cette approche ait déjà été démontré par d'autres études (Jain et al., 2016; Azzari et al., 2017; Lambert et al., 2018), nous remarquons dans notre cas qu'elle est de moindre efficacité pour modéliser les rendements du mil. Une explication probable est liée au fait que la période de pic d'activité des cultures (fin Août/début-mi Septembre) n'est que très faiblement couverte dans notre série temporelle optique, en raison de l'omniprésence des nuages (voir Figure 4.1). Ainsi, les valeurs observées de NDVI sont certainement sous-estimées par rapport à la réalité, ce qui a pu impacter la qualité de la régression linéaire.
Parmi les autres approches évaluées, les modèles d'apprentissage profond (LSTM et CNN) ont obtenu des performances inférieures aux approches traditionnellement utilisées comme les forêts aléatoires ou le perceptron multicouche. Ces résultats sont en contradiction avec la littérature récente (ex. (You et al., 2017; Wolanin et al., 2020; Khaki et al., 2021)), mais semblent liés principalement à notre contexte d'étude. En effet, rappelons que ce dernier est très limité en données de référence et pose dès le départ des défis à l'entraînement supervisé des techniques d'apprentissage profond. Ces méthodes ont surtout montré leur supériorité par rapport aux techniques classiques d'apprentissage automatique dans des scénarios bien différents où les données d'entraînement étaient disponibles à plus large échelle. La disponibilité de données à très petite échelle dans notre cas (66 parcelles suivies au total) freine certainement l'extraction de représentations efficaces permettant aux méthodes d'apprentissage profond de bien généraliser dans cette tâche de modélisation des rendements. Cette éventualité prend d'autant plus de sens que les meilleures performances des modèles LSTM et CNN (R^2 moyen de 0.42), obtenues en employant séparément les groupes de variables explicatives, sont basées sur les indices de végétation que nous pouvons considérer comme des caractéristiques extraites manuellement, plutôt que les données brutes que sont les cœfficients de rétrodiffusion radar ou les bandes spectrales optiques. Toutefois, il convient de mentionner qu'en particulier le CNN a montré du potentiel dans ce contexte qui pose des défis aux techniques d'apprentissage profond. Ceci est effectivement mis en évidence sur les diagrammes de dispersion (Figure 4.3) où de meilleurs accords en faveur du CNN ont été observés entre rendements observés et prédictions moyennées.

Un autre point important abordé dans ce travail est celui relatif à l'uti-

lisation de données multi-sources radar et optique pour la modélisation des rendements du mil. Nous avons tout d'abord employé les variables explicatives multi-sources séparément et obtenu de meilleures performances vis-à-vis de l'utilisation des variables optiques (bandes spectrales ou indices de végétation) par rapport aux cœfficients de rétrodiffusion radar. Ces résultats obtenus pour le mil dans le contexte de la petite agriculture familiale en Afrique subsaharienne sont en contradiction avec d'autres études, bien que sur des systèmes différents, comme celle de Fieuzal and Baup (2017) qui ont pour leur part obtenu une meilleure contribution des cœfficients de rétrodiffusion radar (en polarisation parallèle VV et croisée VH en bande C) par rapport aux bandes spectrales optiques pour la modélisation des rendements du blé. La résolution spatiale des images radar utilisées dans cette étude à savoir TerraSAR-X (3 à 1.5m) et Radarsat-2 (5m) par rapport à Sentinel-1 (10-m) ou encore la canopée qui est plus homogène qu'en petite agriculture familiale caractérisée majoritairement par des systèmes mixtes de culture, sont autant de facteurs pouvant expliquer ces dissimilitudes. Fieuzal et al. (2017) néanmoins, ont obtenu de meilleures performances avec l'utilisation de la réflectance dans les longueurs d'onde du rouge par rapport à celle du signal radar pour l'estimation des rendements du maïs. En somme, il n'est pas étonnant compte tenu des résultats existant dans la littérature que les indications sur les processus biophysiques fournies par l'optique soient plus significatifs dans notre cas que les cœfficients de rétrodiffusion radar. Par la suite, nous avons exploré quatre scénarios possibles de combinaison multi-source pour les différents groupes de variables explicatives. Parmi ces scénarios, seule l'association des cœfficients de rétrodiffusion radar avec des indices de végétation a globalement amélioré la modélisation des rendements. Néanmoins ces améliorations étaient étroitement liées à la méthode de modélisation employée. Étant donné que la complémentarité entre signaux radar et optique est bien établie dans la littérature pour des applications comme la cartographie des cultures ou de l'occupation des sols (Jin et al., 2019; Ienco et al., 2019b), il est inattendu que seules quelques méthodes dans notre étude (RF, CNN) puissent effectivement en tirer parti tandis que le autres (RR, MLP et LSTM) en soient dans l'incapacité. Néanmoins, des observations similaires dans d'autres contextes d'études existent dans la littérature avec des réseaux de neurones artificiels (MLP). Dans l'étude de Fieuzal and Baup (2017), des prévisions de rendements du blé sont conduits tout au long de la saison culturale sur une trentaine de parcelles localisées dans le sud-est de la France, en utilisant des données satellitaires multi-sources radar (TerraSAR-X et Radarsat-2) et optique (Formosat-2 et SPOT-4/5). Parmi diverses configurations testées, incluant des combinaisons entre cœfficients de rétrodiffusion radar et bandes spectrales optiques, c'est plutôt une combi-

naison de cœfficients de rétrodiffusion radar (en polarisation parallèle VV et croisée VH en bande C) et non une combinaison radar et optique qui a fourni les meilleures prévisions en saison. De même, l'étude de Fieuzal et al. (2017), dans un contexte quasi similaire au précédent, a montré de meilleures performances vis-à-vis de la réflectance dans les longueurs d'onde du rouge plutôt qu'une combinaison radar et optique pour la modélisation des rendements du maïs. Le point commun entre ces études et la notre réside dans le fait qu'elles aient été conduites avec des échantillons à très petite échelle. Ainsi, il nous paraît probable que dans le contexte limité en données de référence qui caractérise ces études, toutes les approches de modélisation adoptées ne soient pas pertinentes ou du moins adaptées pour tirer parti des dépendances multi-sources à l'opposé des scénarios en cartographie des cultures ou de l'occupation des sols où la vérité terrain est plus importante et disponible à plus large échelle.

Dans ce travail, nous avons également conduit une phase de prévision des rendements afin de déterminer une période de temps en cours de saison où des prédictions satisfaisantes de rendements pouvaient être obtenues sur le site d'étude. Compte tenu de la proportion de la variabilité des rendements expliquée par notre modèle RF, il apparaît que des prédictions stables et satisfaisantes soient possibles au mieux deux semaines avant la fin de la saison agricole au plus tard en fin Octobre. Des estimations précoces des rendements en cours de saison sont à la clé de stratégies et politiques efficaces en matière de sécurité alimentaire. Notons qu'au Sénégal, les statistiques officielles sur les rendements, obtenues traditionnellement à partir de campagnes sur le terrain usant de stratégies d'échantillonnage stratifiées, coûteuses en temps et en moyens humains, ne sont disponibles que quelques mois après les récoltes (Jacques and Defourny, 2019), typiquement en fin Novembre. Comme montré dans notre étude et précédemment sur d'autres sites (Fieuzal and Baup, 2017), la modélisation des rendements par télédétection est une voie qui peut aider à l'anticipation des récoltes, améliorant ainsi les temps de collecte de statistiques sur la production des cultures, en l'occurrence ici du mil, tout en favorisant une réponse adaptée aux situations de pénurie. Néanmoins, nous ne sous-tendons pas que cette voie devrait remplacer les efforts traditionnels en matière de collecte de données sur le terrain mais plutôt les accompagner et les compléter.

En dernier lieu, nous avons spatialisé les rendements du mil à l'échelle de la parcelle pour les saisons culturales 2018 et 2019. Cette étape finale de cartographie des rendements peut aider à analyser l'hétérogénéité spatiale existante même pour des parcelles contiguës, tout en guidant les pratiques agronomiques et les stratégies visant à minimiser ces disparités. À titre d'exemple, (Jin et al., 2019) ainsi que (Leroux et al., 2020) ont analysé

respectivement récemment pour les cultures de mil et de maïs, les facteurs de la variabilité spatiale des rendements sur base de variables associées à la gestion des parcelles, aux pratiques culturales et aux paramètres édaphiques comme la quantité de nutriments. Dans notre étude, la distribution spatiale des rendements obtenus en fonction des zones d'habitation (anneau de fertilité notamment) semble indiquer que les apports en matière organique sont un déterminant majeur de l'hétérogénéité existante. Par ailleurs, nous avions aussi analysé la distribution spatiale des écarts entre prévisions et estimations ainsi que leur tendances pour les deux saisons culturales considérées. Les prévisions étaient globalement sur-estimées par rapport aux estimations en 2018, notamment dans l'anneau de fertilité tandis qu'en 2019 il s'est produit plutôt l'inverse pour cette zone particulière. Nous n'avons donc pas décelé de tendance claire d'une année à l'autre entre les écarts obtenus pour une zone spécifique. Quelques facteurs liés à la variabilité inter-annuelle des rendements en petite agriculture familiale peuvent expliquer ces dissimilitudes. Par exemple, les effets environnementaux (début de la saison des pluies) ou pratiques agricoles (rotations culturales, cultures mixtes ou intercalées, quantité de fertilisants appliqués) sont autant de facteurs qui causent de l'hétérogénéité entre les récoltes des saisons culturales successives. Par conséquent, il peut être délicat d'obtenir des tendances similaires entre les années en considérant les mêmes fenêtres temporelles pour prévoir les rendements. Pour finir, rappelons que toute l'analyse qui résulte de la spatialisation des rendements du mil est sujette à la qualité de la cartographie de l'occupation du sol réalisée en amont et que des erreurs de classification peuvent entraîner de ce fait de possibles biais dans le raisonnement qui s'en est suivi.

En fin de compte, soulignons tout de même que notre étude est conduite à l'échelle d'une petite zone d'étude (17 km^2) et est restreinte en terme d'échantillons observés. La difficulté à collecter des données de référence à plus large échelle pour des cultures vivrières comme le mil, principalement en raison des coûts d'acquisition, est l'un des facteurs majeurs qui empêchent le passage à l'échelle de nombreuses recherches sur le suivi des rendements en petite agriculture familiale notamment en Afrique subsaharienne.

4.4 Conclusion

Dans ce chapitre, nous nous sommes intéressés à la modélisation des rendements dans le cas de la petite agriculture familiale en Afrique subsaharienne, qui représente un enjeu majeur pour la sécurité alimentaire dans la sous-région. Bien que la disponibilité d'images de télédétection multi-sources et à hautes résolutions spatiale et temporelle soit un atout pour le suivi des

rendements en petite agriculture familiale, très peu d'études se sont penchées sur la combinaison de données radar et optique dans ce contexte. De plus, les techniques d'apprentissage profond, bien que omniprésentes dans d'autres cas d'étude, ont reçu très peu d'attention pour la modélisation des rendements en petite agriculture familiale, caractérisée par des données de référence très limitées. Ce chapitre de la thèse a ainsi traité de l'estimation (en fin de saison) et de la prévision (en cours de saison) des rendements du mil sur le site du bassin arachidier au Sénégal. Nous utilisons à cette fin des séries temporelles d'images radar et optique (Sentinel-1 et Sentinel-2) et un large panel d'approches comprenant à la fois des techniques usuellement adoptées pour la modélisation des rendements et des modèles d'apprentissage profond. Parmi les approches évaluées en phase d'estimation, la méthode des forêts aléatoires a expliqué la plus grande part de la variabilité des rendements observés sur le site d'étude, tandis que celle des réseaux de neurones convolutifs a montré du potentiel dans ce contexte qui pose des défis aux techniques d'apprentissage profond. La combinaison des données multi-source, et plus précisément celle des cœfficients de rétrodiffusion radar avec des indices de végétation, a pour sa part été concluante en améliorant les performances de modélisation des rendements. Cependant, les améliorations observées étaient étroitement liées à l'approche de modélisation adoptée en l'occurrence les forêts aléatoires et réseaux de neurones convolutifs. Enfin, la prévision des rendements en cours de saison culturale a révélé que des prédictions stables et satisfaisantes pouvaient être obtenues au mieux deux semaines avant la fin de la saison agricole. Bien que ces prévisions soient très rapprochées de la période de récolte, elles peuvent néanmoins contribuer à accélérer le recueil des informations sur la production agricole en accompagnant et en complétant les efforts de collecte sur le terrain. Dans l'ensemble, nos résultats montrent une certaine plus-value de la combinaison multi-source pour la modélisation des rendements en petite agriculture familiale et consolident l'utilisation des forêts aléatoires usuellement adoptée dans ce contexte. Par ailleurs, ce contexte actuellement limité en données de référence, ne semble pas encore, au vu de notre évaluation, tout à fait adapté à la modélisation des rendements agricoles à partir de l'entraînement supervisé, de bout en bout tout du moins, des techniques d'apprentissage profond. Ainsi, d'autres stratégies d'entraînement telles que l'apprentissage semi-supervisé ou l'apprentissage auto-supervisé, qui requièrent moins de données de référence, pourraient permettre à ces techniques, et notamment au CNN qui a montré du potentiel dans notre évaluation, de devenir à leur tour des standards pour la petite agriculture familiale.

Conclusion et Perspectives

Contributions de la thèse

L'objectif général de cette thèse est de proposer des contributions méthodologiques pour améliorer l'évaluation de la production agricole dans les systèmes de suivi des cultures à partir de la télédétection multi-source et des techniques d'apprentissage profond. L'évaluation de la production agricole repose sur le suivi des surfaces cultivées et des rendements des cultures. Ainsi, nous avons mené des travaux relatifs à ces deux aspects en ayant en outre un regard croisé vis-à-vis de l'application de nos méthodes sur des sites d'étude contrastés, localisés à la fois en agriculture conventionnelle et en petite agriculture familiale.

Nous avons proposé deux approches méthodologiques pour caractériser l'occupation du sol et identifier les surfaces cultivées. Notre première approche est basée sur des RNNs équipés de mécanismes d'attention. Cette approche est mise au point pour le couplage de séries temporelles multi-sources radar et optique (Sentinel-1 et Sentinel-2) et incorpore dans le processus, des connaissances spécifiques de domaine à travers une stratégie de pré-entraînement. Dans le but d'intégrer davantage de données satellitaires multi-sources, nous avons par la suite proposé une seconde approche reposant sur des CNNs. Cette approche combine à la fois des informations satellitaires multi-sources et multi-temporelles mais également multi-échelles en couplant des séries temporelles radar et optique (Sentinel-1 et Sentinel-2) à haute résolution ainsi qu'une image optique à très haute résolution spatiale (SPOT). Chacune de ces approches est évaluée à des échelles territoriale et locale en prenant en compte de sites d'études contrastés par leurs paysages, conditions agro-climatiques et pratiques culturales. De manière générale, l'évaluation des approches proposées a montré que les techniques d'apprentissage profond semblent mieux adaptés que les méthodes courantes d'apprentissage automatique pour tirer parti de la complémentarité des données multi-sources, multi-temporelles et multi-échelles, à mesure qu'il y ait une quantité suffisante d'échantillons étiquetés pour leur entraînement. Les modèles d'ap-

prentissage profond employés dans cette thèse à savoir les RNNs et CNNs ont leurs avantages et inconvénients respectifs. Les RNNs sont par exemple, par rapport aux CNNs, généralement plus lents en phase d'entraînement tandis qu'ils ont l'avantage de pouvoir traiter des séquences de tailles variables. Néanmoins, au delà du couplage multi-source par un réseau de neurones profond, nos travaux ont aussi montré que d'autres composantes méthodologiques méritent d'être prises en compte. Soulignons par exemple l'intégration de classifieurs auxiliaires dans le processus de combinaison multi-source que nous avons étendu ultérieurement pour distiller les connaissances entre couches plus profondes et moins profondes. De même, notons l'intégration de connaissances spécifiques de domaine, à savoir les relations hiérarchiques entre classes d'occupation du sol, qui semble bénéfique au processus d'entraînement des réseaux de neurones. Si peu d'études à l'époque de ces travaux avaient considéré l'intégration de connaissances à priori dans l'entraînement des réseaux de neurones, de plus en plus d'études présentement semblent s'orienter en ce sens (Bertinetto et al., 2020; Garnot and Landrieu, 2020; Turkoglu et al., 2021). Pour finir, les résultats obtenus nous incitent également à nous questionner sur l'apport éventuel de l'intégration d'autres sources de données pouvant être par exemple un modèle numérique de surface.

Nous avons également contribué dans cette thèse à l'évaluation des rendements des surfaces cultivées en proposant une démarche méthodologique comparative appliquée à l'échelle locale de la petite agriculture familiale en Afrique subsaharienne. L'évaluation des rendements dans cette région représente à l'heure actuelle un enjeu majeur de sécurité alimentaire tandis que peu de données sur les rendements observés sont disponibles. Dans ce contexte inéluctablement limité en données de référence, nous avons traité de l'estimation (en fin de saison) et de la prévision (en cours de saison) des rendements d'une culture de base dans la région, le mil, principalement destiné à la consommation des ménages. En employant des variables explicatives issues de séries temporelles multi-sources radar et optique (Sentinel-1 et Sentinel-2), nous avons donc évalué les performances de modèles d'apprentissage profond notamment LSTM et CNN par rapport à des approches traditionnellement utilisées comme les forêts aléatoires. Le travail d'investigation réalisé dans ce cas d'étude précis n'a pas montré de plus-value manifeste à l'emploi de techniques d'apprentissage profond, leurs performances restant globalement inférieures à celles des forêts aléatoires, une approche courante pour la modélisation des rendements. Également, le couplage de données multi-sources et multi-temporelles n'a été bénéfique que pour une poignée d'approches évaluées (forêts aléatoires et CNN). Il y a certes du potentiel à tirer des méthodes d'apprentissage profond notamment le CNN dans le contexte actuel de la petite agriculture familiale, mais cette dernière ne semble pas adaptée à

l'entraînement supervisé de bout en bout des méthodes d'apprentissage profond. À défaut de ne pouvoir collecter une quantité suffisante de données sur les rendements observés dans le cas de la petite agriculture familiale, d'autres types d'apprentissage (semi-supervisé, auto-supervisé, par transfert) mériteraient d'être examinés (van Engelen and Hoos, 2020; Jing and Tian, 2020; Kaneko et al., 2017).

Perspectives aux travaux réalisés

Les contributions méthodologiques apportées dans cette thèse sont évaluées à des échelles territoriale et locale. À l'égard des systèmes de suivi des cultures, leur intégration devra être synonyme d'un passage à plus large échelle à savoir régionale, nationale voire globale. Au vu des progrès réalisés en termes de capacité de calcul et mémoire des ordinateurs, le passage à l'échelle d'un point de vue computationnel est envisageable sans obstacle majeur. Néanmoins, certaines entraves peuvent par exemple provenir de modalités manquantes dans l'approche multi-source. En effet, il peut survenir qu'une ou plusieurs sources de données, par exemple des images optiques à très haute résolution spatiale, ne soient disponibles que sur une portion des sites d'études en raison de coûts d'acquisition élevés, d'une fauchée souvent réduite par rapport aux images de résolution spatiale moins fine ou encore à cause de la nébulosité fréquente. Au lieu d'exclure ces données manquantes, une piste pourrait se trouver dans la distillation des connaissances (Hinton et al., 2015) et l'hallucination de modalités (Hoffman et al., 2016). La distillation de connaissances a été proposée initialement pour transférer les connaissances d'un grand modèle aussi appelé enseignant, vers un modèle plus petit désigné étudiant, en faisant mimer la sortie du modèle enseignant au modèle étudiant. L'enseignant serait dans notre cas de figure, un réseaux de neurones entraîné avec l'ensemble des modalités (non manquantes et manquantes en partie) tandis que l'étudiant serait le réseau entraîné uniquement avec les modalités non manquantes. Le modèle étudiant permettrait dès lors de pallier les sources manquantes au moment de l'inférence. L'hallucination de modalités s'inspire de la distillation de connaissances entre réseaux de neurones. Elle permet de simuler la présence de la modalité en partie manquante pendant la phase d'inférence. Pour ce faire, un premier modèle est entraîné uniquement avec les modalités non manquantes et un second uniquement avec les modalités manquantes en partie. Le premier modèle est doté d'un réseau ou branche d'hallucination lui permettant d'apprendre une représentation associée aux modalités en partie manquantes à partir du second modèle. C'est donc ce modèle entraîné avec les modalités non manquantes mais dis-

posant de connaissances provenant des modalités en partie manquantes qui est déployé en phase d'inférence.

Également, d'autres obstacles liés aux données étiquetées en classification de l'occupation du sol peuvent entraver un passage à l'échelle vis-à-vis des systèmes de suivi des cultures. Ils proviennent par exemple du fait que les données étiquetées ne soient pas toujours disponibles sur l'entièreté des zones d'intérêt. Cependant, l'occupation du sol tend à différer d'une région à l'autre. Ainsi, disposer de données étiquetées rien qu'à l'échelle d'une région n'aide donc pas nécessairement dans la cartographie précise de l'autre. Tout autant, la qualité des étiquettes est extrêmement variable d'une région à l'autre compte tenu notamment de procédures de collecte diverses. Tout ceci pose donc des problèmes vis-à-vis de l'entraînement des modèles et de leur déploiement en phase d'inférence. D'autre part, une situation fréquemment rencontrée est celle de l'actualisation des étiquettes. En raison des coûts liés à la collecte des données, les étiquettes disponibles ne reflètent pas toujours l'occupation du sol de la période d'intérêt. Dès lors, des pistes à explorer face à ces obstacles pourraient résider dans l'apprentissage par transfert et plus précisément l'adaptation de domaine (Kouw, 2018; Kouw and Loog, 2021). L'adaptation de domaine est un champ en apprentissage automatique qui traite de données dont la distribution n'est pas stationnaire dans l'espace et le temps (Tuia et al., 2016). Il fait intervenir un domaine source où les données de référence sont généralement disponibles, de bonne qualité et à jour ainsi qu'un domaine cible, celui où les annotations sont rares voire inexistantes. L'adaptation de domaine peut être ainsi explorée pour examiner les possibilités de transfert spatial (Lucas et al., 2020a) d'un modèle entraîné sur une région particulière (domaine source) à une autre (domaine cible) de même que celles du transfert temporel d'un modèle entraîné sur une même région mais à des périodes de temps différentes. Si des contributions récentes dans un cadre d'utilisation de données mono-source ont été apportées à ce sujet (ex. Lucas et al. (2020b, 2021)), le contexte multi-source pose plus de défis et est toujours ouvert à investigation (Munro and Damen, 2020).

Enfin, un autre obstacle au passage à l'échelle auquel nous nous sommes également confrontés dans cette thèse (exemple du site du bassin arachidier au Sénégal), est celui lié à la faible quantité des données étiquetées disponible sur certaines zones d'intérêt. Ainsi, que ce soit pour des tâches de classification ou de régression, il est probable d'envisager d'autres modes d'apprentissage comme l'apprentissage semi-supervisé ou auto-supervisé (self-supervised). Comme évoqué précédemment en Section 1.2, l'approche semi-supervisée permet de tirer à la fois parti de données étiquetées disponibles en faible quantité et de données non étiquetées souvent disponibles largement en plus grand ensemble. Dans l'analyse des images satellitaires, un nombre

considérable de données non annotées de même distribution que les données annotées, est disponible en phase d'entraînement et peut être mis à contribution. Partant toujours du principe de la disponibilité de ce large ensemble de données non annotées, une autre direction pourrait être celle de l'apprentissage auto-supervisé. Ce mode d'apprentissage a émergé avec l'apprentissage profond et est devenu un standard en traitement automatique du langage avec les travaux comme Word2Vec (Mikolov et al., 2013) ou BERT (Devlin et al., 2019). L'apprentissage auto-supervisé d'un réseau de neurones consiste à faire précéder sa tâche principale d'apprentissage d'une tâche de prétexte dans le but de lui faire apprendre des représentations utiles pour la tâche cible. Cette approche permet ainsi d'obtenir de meilleures performances sur la tâche d'apprentissage principale comparé à un apprentissage supervisé effectué de bout-en-bout à partir de la faible quantité de données annotées. L'apprentissage auto-supervisé commence à émerger également dans l'analyse d'images satellitaires notamment mono-source (ex. Yuan and Lin (2021)) où il est possible de construire des tâches de prétexte à partir du large ensemble de données non annotées existant. Ce champ reste également ouvert à investigation dans un contexte de multiples modalités.

Références

Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., and Süsstrunk, S. (2012). SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(11) :2274–2282. 104

Audebert, N., Le Saux, B., and Lefèvre, S. (2017). Semantic segmentation of earth observation data using multimodal and multi-scale deep networks. In *Computer Vision – ACCV 2016*, pages 180–196. Springer International Publishing. 42

Azzari, G., Jain, M., and Lobell, D. (2017). Towards fine resolution global maps of crop yields : Testing multiple methods and satellites in three countries. *Remote Sensing of Environment*, page In press. 100, 113

Bahdanau, D., Cho, K., and Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. In Bengio, Y. and LeCun, Y., editors, *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*. 23, 25, 68

Benedetti, P., Ienco, D., Gaetano, R., Ose, K., Pensa, R. G., and Dupuy, S. (2018). M3 fusion : A deep learning architecture for multiscale multimodal multitemporal satellite data fusion. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(12) :4939–4949. 77

Berger, M., Moreno, J., Johannessen, J. A., Levelt, P. F., and Hanssen, R. F. (2012). Esa's sentinel missions in support of earth system science. *Remote Sensing of Environment*, 120 :84 – 90. xviii

Bertinetto, L., Mueller, R., Tertikas, K., Samangooei, S., and Lord, N. A. (2020). Making better mistakes : Leveraging class hierarchies with deep networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 120

Blaschke, T. (2010). Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65(1) :2–16. 42

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1) :5–32. 11, 100

Britz, D., Guan, M. Y., and Luong, M. (2017). Efficient attention using a fixed-size memory representation. In *EMNLP*, pages 392–400. 25, 68

Cho, K., van Merriënboer, B., Bahdanau, D., and Bengio, Y. (2014). On the properties of neural machine translation : Encoder–decoder approaches. In *Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation*, pages 103–111. Association for Computational Linguistics. 22

Choi, H., Cho, K., and Bengio, Y. (2018). Fine-grained attention mechanism for neural machine translation. *Neurocomputing*, 284 :171–176. 68

Chung, J., Gulcehre, C., Cho, K., and Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. In *NIPS 2014 Workshop on Deep Learning, December 2014*. 22

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1) :37–46. 27

Cresson, R. and Saint-Geours, N. (2015). Natural color satellite image mosaicking using quadratic programming in decorrelated color space. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(8) :4151–4162. 75

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). BERT : Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics : Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics. 25, 123

Di Gregorio, A. (2005). *Land cover classification system : classification concepts and user manual : LCCS*, volume 2. Food & Agriculture Org. 43

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., and Bargellini, P. (2012). Sentinel-2 : ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sensing of Environment*, 120 :25–36. 3

Dupuy, S., Gaetano, R., and Mézo, L. L. (2020). Mapping land cover on reunion island in 2017 using satellite imagery and geospatial ground data. *Data in Brief*, 28 :104934. 32

Erinjery, J., Singh, M., and Kent, R. (2018). Mapping and assessment of vegetation types in the tropical rainforests of the western ghats using multispectral sentinel-2 and sar sentinel-1 satellite imagery. *Remote Sensing of Environment*, 216 :345–354. 54

Félix, G., Diedhiou, I., Le Garff, M., Timmermann, C., Clermont-Dauphin, C., Cournac, L., Groot, J., and Tittonell, P. (2018). Use and management of biodiversity by smallholder farmers in semi-arid West Africa. *Global Food Security*, 18. 99

Fieuzal, R. and Baup, F. (2017). Forecast of wheat yield throughout the agricultural season using optical and radar satellite images. *International Journal of Applied Earth Observation and Geoinformation*, 59 :147 – 156. 97, 114, 115

Fieuzal, R., Marais-Sicre, C., and Baup, F. (2017). Estimation of corn yield using multi-temporal optical and radar satellite data and artificial neural networks. *International Journal of Applied Earth Observation and Geoinformation*, 57 :14 – 23. 97, 114, 115

Fritz, S., See, L., McCallum, I., Schill, C., Obersteiner, M., van der Velde, M., Boettcher, H., Havlík, P., and Achard, F. (2011). Highlighting continued uncertainty in global land cover maps for the user community. *Environmental Research Letters*, 6(4) :044005. xvii

Fritz, S., See, L., Mccallum, I. a. N., You, L., Bun, A., Moltchanova, E., Duerauer, M., Albrecht, F., Schill, C., Perger, C., Havlik, P., Mosnier, A., Thornton, P., Wood-sichra, U., Herrero, M., and Becker-Reshef, I. (2015). Mapping global cropland and field size. *Global Change Biology*. xvii

Gaetano, R., Ienco, D., Ose, K., and Cresson, R. (2018). A two-branch CNN architecture for land cover classification of PAN and MS imagery. *Remote. Sens.*, 10(11) :1746. 73

Gao, B. (1996). Ndwi—a normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment*, 58(3) :257 – 266. 7

Garnot, V. S. F. and Landrieu, L. (2020). Metric-guided prototype learning. *CoRR*, abs/2007.03047. 120

Garnot, V. S. F. and Landrieu, L. (2021). Panoptic segmentation of satellite image time series with convolutional temporal attention networks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 4872–4881. 42

Gers, F. (2001). *Long Short-Term Memory in Recurrent Neural Networks*. PhD thesis, École Polytechnique Fédérale de Lausanne. 20

Gitelson, A. A., Gritz, Y., and Merzlyak, M. N. (2003). Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *Journal of Plant Physiology*, 160(3) :271 – 282. 7

Glorot, X., Bordes, A., and Bengio, Y. (2011). Deep sparse rectifier neural networks. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, AISTATS 2011, Fort Lauderdale, USA, April 11-13, 2011*, volume 15 of *JMLR Proceedings*, pages 315–323. JMLR.org. 16

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press. 12, 15, 18

Gou, J., Yu, B., Maybank, S. J., and Tao, D. (2020). Knowledge distillation : A survey. *CoRR*, abs/2006.05525. 72, 82

Hinton, G., Vinyals, O., and Dean, J. (2015). Distilling the knowledge in a neural network. In *NIPS Deep Learning and Representation Learning Workshop*. 72, 121

Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. R. (2012). Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv :1207.0580*. 18

Hochreiter, S. and Schmidhuber, J. (1996). LSTM can solve hard long time lag problems. In *NIPS*, pages 473–479. 20

Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression : Biased estimation for nonorthogonal problems. *Technometrics*, 12(1) :55–67. 100

Hoffman, J., Gupta, S., and Darrell, T. (2016). Learning with side information through modality hallucination. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, pages 826–834. IEEE Computer Society. 121

Hong, D., Gao, L., Yokoya, N., Yao, J., Chanussot, J., Du, Q., and Zhang, B. (2020). More diverse means better : Multimodal deep learning meets remote-sensing imagery classification. *IEEE Transactions on Geoscience and Remote Sensing*, pages 1–15. 77

Hou, S., Liu, X., and Wang, Z. (2017). Dualnet : Learn complementary features for image recognition. In *ICCV*, pages 502–510. 49

Hu, Y., Soltoggio, A., Lock, R., and Carter, S. (2019). A fully convolutional two-stream fusion network for interactive image segmentation. *Neural Networks*, 109 :31–42. 72

Ienco, D., Gaetano, R., Dupaquier, C., and Maurel, P. (2017). Land Cover Classification via Multitemporal Spatial Data by Deep Recurrent Neural Networks. *IEEE Geoscience and Remote Sensing Letters*, 14(10) :1685–1689. xix

Ienco, D., Gaetano, R., Interdonato, R., Ose, K., and Minh, D. H. T. (2019a). Combining sentinel-1 and sentinel-2 time series via RNN for object-based land cover classification. In *2019 IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2019, Yokohama, Japan, July 28 - August 2, 2019*, pages 4881–4884. IEEE. 54, 55

Ienco, D., Interdonato, R., Gaetano, R., and Minh, D. H. T. (2019b). Combining sentinel-1 and sentinel-2 satellite image time series for land cover mapping via a multi-source deep learning architecture. *ISPRS Journal of Photogrammetry and Remote Sensing*, 158 :11 – 22. 49, 58, 114

Inglada, J., Vincent, A., Arias, M., Tardy, B., Morin, D., and Rodes, I. (2017). Operational high resolution land cover map production at the country scale using satellite image time series. *Remote Sensing*, 9(1) :95. 40

Interdonato, R., Ienco, D., Gaetano, R., and Ose, K. (2019). Duplo : A dual view point deep learning architecture for time series classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 149 :91 – 104. 49, 77

Ioffe, S. and Szegedy, C. (2015). Batch normalization : Accelerating deep network training by reducing internal covariate shift. In *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6-11 July 2015*, volume 37, pages 448–456. 18

Jacques, D. C. and Defourny, P. (2019). Accuracy requirements for early estimation of crop production in senegal. 115

Jain, M., Srivastava, A., Balwinder-Singh, Joon, R., McDonald, A., Royal, K., Lisaius, M., and Lobell, D. (2016). Mapping Smallholder Wheat Yields and Sowing Dates Using Micro-Satellite Data. *Remote Sensing*, 8(10) :860. 97, 100, 113

Ji, S., Zhang, C., Xu, A., Shi, Y., and Duan, Y. (2018). 3d convolutional neural networks for crop classification with multi-temporal remote sensing images. *Remote Sensing*, 10(2) :75. 76

Jiang, Z., Huete, A. R., Didan, K., and Miura, T. (2008). Development of a two-band enhanced vegetation index without a blue band. *Remote Sensing of Environment*, 112(10) :3833 – 3845. 7

Jin, Z., Azzari, G., Burke, M., Aston, S., and Lobell, D. B. (2017). Mapping smallholder yield heterogeneity at multiple scales in eastern africa. *Remote Sensing*, 9(9) :931. 97

Jin, Z., Azzari, G., You, C., Di Tommaso, S., Aston, S., Burke, M., and Lobell, D. B. (2019). Smallholder maize area and yield mapping at national scales with google earth engine. *Remote sensing of environment*, 228 :115–128. 96, 114, 115

Jing, L. and Tian, Y. (2020). Self-supervised visual feature learning with deep neural networks : A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–1. 121

Kamilaris, A. and Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture : A survey. *Computers and Electronics in Agriculture*, 147(July 2017) :70–90. xix

Kamir, E., Waldner, F., and Hochman, Z. (2020). Estimating wheat yields in australia using climate records, satellite image time series and machine learning methods. *ISPRS Journal of Photogrammetry and Remote Sensing*, 160 :124 – 135. 97

Kaneko, A., Kennedy, T., Mei, L., Sintek, C., Burke, M., Ermon, S., and Lobell, D. (2017). Deep Learning For Crop Yield Prediction in Africa. *International Conference on Machine Learning AI for Social Good Workshop, Long Beach, United States, 2019*, pages 1–5. 97, 121

Karamanolakis, G., Hsu, D., and Gravano, L. (2019). Weakly supervised attention networks for fine-grained opinion mining and public health. In *W-NUT@EMNLP*, pages 1–10. 47

Khaki, S., Pham, H., and Wang, L. (2021). Simultaneous corn and soybean yield prediction from remote sensing data using deep transfer learning. *Scientific Reports*, 11(1) :11132. 97, 113

Khaki, S. and Wang, L. (2019). Crop yield prediction using deep neural networks. *Frontiers in Plant Science*, 10. xix, 97

Khaki, S., Wang, L., and Archontoulis, S. V. (2020). A CNN-RNN Framework for Crop Yield Prediction. *Frontiers in Plant Science*, 10(January) :1–14. 97

Kim, N., Ha, K.-J., Park, N.-W., Cho, J., Hong, S., and Lee, Y.-W. (2019). A comparison between major artificial intelligence models for crop yield prediction : Case study of the midwestern united states, 2006–2015. *ISPRS International Journal of Geo-Information*, 8(5) :240. 97

Kingma, D. P. and Ba, J. (2015). Adam : A method for stochastic optimization. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*. 57, 78, 100

Kogan, F., Kussul, N., Adamenko, T., Skakun, S., Kravchenko, O., Kryvobok, O., Shelestov, A., Kolotii, A., Kussul, O., and Lavrenyuk, A. (2013). Winter wheat yield forecasting in ukraine based on earth observation, meteorological data and biophysical models. *International Journal of Applied Earth Observation and Geoinformation*, 23 :192–203. 96

Kouw, W. M. (2018). An introduction to domain adaptation and transfer learning. *ArXiv*, abs/1812.11806. 122

Kouw, W. M. and Loog, M. (2021). A review of domain adaptation without target labels. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(3) :766–785. 122

Kussul, N., Lavreniuk, M., Skakun, S., and Shelestov, A. (2017). Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data. *IEEE Geoscience and Remote Sensing Letters*, 14(5) :778–782. xix

Lambert, M.-J., Traoré, P., Blaes, X., Baret, P., and Defourny, P. (2018). Estimating smallholder crops production at village level from Sentinel-2 time series in Mali's cotton belt. *Remote Sensing of Environment*, 216. 96, 100, 113

Lassalle, P., Inglada, J., Michel, J., Grizonnet, M., and Malik, J. (2015). A scalable tile-based framework for region-merging segmentation. *IEEE Transactions on Geoscience and Remote Sensing*, 53(10) :5473–5485. 51

Lecun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521(7553) :436–444. xix, 11

LeCun, Y., Boser, B. E., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W. E., and Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. *Neural Computation*, 1(4) :541–551. 12

Leroux, L., Castets, M., Baron, C., Escorihuela, M.-J., Bégué, A., and Seen, D. L. (2019). Maize yield estimation in west africa from crop process-induced combinations of multi-domain remote sensing indices. *European Journal of Agronomy*, 108 :11–26. 96

Leroux, L., Falconnier, G., Diouf, A., Ndao, B., Gbodjo, J., Tall, L., Balde, A., Clermont-Dauphin, C., Bégué, A., Affholder, F., and Roupsard, O. (2020). Using remote sensing to assess the effect of trees on millet yield in complex parklands of central senegal. *Agricultural Systems*, 184 :102918. 37, 97, 115

Leroux, L., Jolivot, A., Bégué, A., Lo Seen, D., and Zoungrana, B. (2014). How Reliable is the MODIS Land Cover Product for Crop Mapping Sub-Saharan Agricultural Landscapes ? *Remote Sensing*, 6 :8541–8564. xviii

Lobell, D. B. (2013). The use of satellite data for crop yield gap analysis. *Field Crops Research*, 143 :56 – 64. 97

Lobell, D. B., Di Tommaso, S., You, C., Yacoubou Djima, I., Burke, M., and Kilic, T. (2019). Sight for sorghums : Comparisons of satellite- and ground-based sorghum yield estimates in mali. *Remote Sensing*, 12(1) :100. 97

Lucas, B., Pelletier, C., Schmidt, D., Webb, G. I., and Petitjean, F. (2020a). Unsupervised domain adaptation techniques for classification of satellite image time series. In *IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium*, pages 1074–1077. 122

Lucas, B., Pelletier, C., Schmidt, D., Webb, G. I., and Petitjean, F. (2021). A Bayesian-inspired, deep learning-based, semi-supervised domain adaptation technique for land cover mapping. *Machine Learning*. 122

Lucas, B., Pelletier, C., Schmidt, D. F., Webb, G. I., and Petitjean, F. (2020b). Unsupervised domain adaptation techniques for classification of satellite image time series. In *IGARSS*, pages 1074–1077. IEEE. 122

Luong, M., Pham, H., and Manning, C. D. (2015). Effective approaches to attention-based neural machine translation. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, EMNLP 2015*, pages 1412–1421. 24

Ma, L., Schmitt, M., and Zhu, X. (2020). Uncertainty analysis of object-based land-cover classification using sentinel-2 time-series data. *Remote Sensing*, 12(22) :3798. 93

Marteau, R., Sultan, B., Moron, V., Alhassane, A., Baron, C., and Traoré, S. B. (2011). The onset of the rainy season and farmers' sowing strategy for pearl millet cultivation in southwest niger. *Agricultural and Forest Meteorology*, 151(10) :1356 – 1369. 37

Maxwell, A. E., Warner, T. A., and Fang, F. (2018). Implementation of machine-learning classification in remote sensing : an applied review. *International Journal of Remote Sensing*, 39(9) :2784–2817. 65

Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013). Efficient estimation of word representations in vector space. In Bengio, Y. and LeCun, Y., editors, *1st International Conference on Learning Representations, ICLR 2013, Scottsdale, Arizona, USA, May 2-4, 2013, Workshop Track Proceedings*. 123

Minsky, M. and Papert, S. (1969). An introduction to computational geometry. *Cambridge tiass., HIT*. 12

Munro, J. and Damen, D. (2020). Multi-modal domain adaptation for fine-grained action recognition. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, WA, USA, June 13-19, 2020*, pages 119–129. IEEE. 122

Pelletier, C., Webb, G., and Petitjean, F. (2019). Temporal convolutional neural network for the classification of satellite image time series. *Remote Sensing*, 11(5) :523. 43, 54, 55, 73

Pérez-Hoyos, A., Rembold, F., Kerdiles, H., and Gallego, J. (2017). Comparison of global land cover datasets for cropland monitoring. *Remote Sensing*, 9(11). xvii

Qi, J., Chehbouni, A., Huete, A., Kerr, Y., and Sorooshian, S. (1994). A modified soil adjusted vegetation index. *Remote Sensing of Environment*, 48(2) :119 – 126. 7

Quegan, S. and Yu, J. J. (2001). Filtering of multichannel sar images. *IEEE Transactions on Geoscience and Remote Sensing*, 39(11) :2373–2379. 39

Rembold, F., Atzberger, C., Savin, I., and Rojas, O. (2013). Using low resolution satellite imagery for yield prediction and yield anomaly detection. remote sens. 2013, 5, 1704-1733. *Remote Sensing*, 5(11). xviii, 96, 97, 100

Rembold, F., Meroni, M., Urbano, F., Csak, G., Kerdiles, H., Perez-Hoyos, A., Lemoine, G., Leo, O., and Negre, T. (2019). Asap : A new global early warning system to detect anomaly hot spots of agricultural production for food security analysis. *Agricultural Systems*, 168 :247–257. xvii, xviii

Rembold, F., Meroni, M., Urbano, F., Lemoine, G., Kerdiles, H., Perez-Hoyos, A., and Csak, G. (2017). Asap - anomaly hot spots of agricultural production, a new early warning decision support system developed by the joint research centre. In *2017 9th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp)*, pages 1–5. xvii

Rosenblatt, F. (1958). The perceptron : A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6) :386–408. 12

Rouse, J. W., Hass, R. H., Schell, J., and Deering, D. (1973). Monitoring vegetation systems in the great plains with ERTS. *Third Earth Resources Technology Satellite (ERTS) symposium*, 1 :309–317. 7

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088) :533–536. 12, 100

Rußwurm, M. and Körner, M. (2018). Multi-temporal land cover classification with sequential recurrent encoders. *ISPRS Int. J. Geo-Information*, 7(4) :129. xix

Sakamoto, T., Gitelson, A. A., and Arkebauer, T. J. (2013). Modis-based corn grain yield estimation model incorporating crop phenology information. *Remote Sensing of Environment*, 131 :215–231. 96

Schmidhuber, J. (2015). Deep learning in neural networks : An overview. *Neural Networks*, 61 :85–117. xix, 11

Schmitt, M. and Zhu, X. X. (2016). Data fusion and remote sensing : An ever-growing relationship. *IEEE Geoscience and Remote Sensing Magazine*, 4(4) :6–23. xviii

Sirko, W., Kashubin, S., Ritter, M., Annkah, A., Bouchareb, Y. S. E., Dauphin, Y. N., Keysers, D., Neumann, M., Cissé, M., and Quinn, J. (2021). Continental-scale building detection from high resolution satellite imagery. *CoRR*, abs/2107.12283. 42

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout : A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(56) :1929–1958. 18

Sutskever, I., Vinyals, O., and Le, Q. V. (2014). Sequence to sequence learning with neural networks. In *Advances in Neural Information Processing Systems 27 : Annual Conference on Neural Information Processing Systems 2014*, pages 3104–3112. 23

Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, 8(2) :127 – 150. xviii, 7, 97

Tuia, D., Persello, C., and Bruzzone, L. (2016). Domain adaptation for the classification of remote sensing data : An overview of recent advances. *IEEE Geoscience and Remote Sensing Magazine*, 4(2) :41–57. 122

Turkoglu, M. O., D'Aronco, S., Perich, G., Liebisch, F., Streit, C., Schindler, K., and Wegner, J. D. (2021). Crop mapping from image time series : deep learning with multi-scale label hierarchies. *CoRR*, abs/2102.08820. 120

United Nations (2015). Transforming our world : The 2030 agenda for sustainable development. https://sustainabledevelopment.un.org/post2015/transformingourworld/publication. En ligne ; Consulté le 26 Juin 2021. xvii

Valero, S., Arnaud, L., Planells, M., Ceschia, E., and Dedieu, G. (2019). Sentinel's classifier fusion system for seasonal crop mapping. In *2019 IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2019, Yokohama, Japan, July 28 - August 2, 2019*, pages 6243–6246. IEEE. 54

van der Maaten, L. and Hinton, G. (2008). Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(86) :2579–2605. 87

van Engelen, J. E. and Hoos, H. H. (2020). A survey on semi-supervised learning. *Machine Learning*, 109(2) :373–440. 121

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems 30 : Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, pages 5998–6008. 25

Veloso, A., Mermoz, S., Bouvet, A., Le Toan, T., Planells, M., Dejoux, J., and Ceschia, E. (2017). Understanding the temporal behavior of crops using Sentinel-1 and Sentinel-2-like data for agricultural applications. *Remote Sensing of Environment*, 199 :415–426. xviii, 96

Waldner, F. and Defourny, P. (2017). Where can pixel counting area estimates meet user-defined accuracy requirements ? *International Journal of Applied Earth Observation and Geoinformation*, 60 :1–10. xviii

Waldner, F., Fritz, S., Di Gregorio, A., and Defourny, P. (2015). Mapping priorities to focus cropland mapping activities : Fitness assessment of existing global, regional and national cropland maps. *Remote Sensing*, 7(6) :7959–7986. xvii

Waldner, F., Fritz, S., Di Gregorio, A., Plotnikov, D., Bartalev, S., Kussul, N., Gong, P., Thenkabail, P., Hazeu, G., Klein, I., Löw, F., Miettinen, J., Dadhwal, V. K., Lamarche, C., Bontemps, S., and Defourny, P. (2016). A unified cropland layer at 250 m for global agriculture monitoring. *Data*, 1(1). xvii

Wang, L. and Yoon, K. J. (2021). Knowledge distillation and student-teacher learning for visual intelligence : A review and new outlooks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–1. 72, 82

Wang, P., Zhang, H., and Patel, V. M. (2017). Sar image despeckling using a convolutional neural network. *IEEE Signal Processing Letters*, 24(12) :1763–1767. 79

Werbos, P. J. (1974). *Beyond Regression : New Tools for Prediction and Analysis in the Behavioral Sciences*. PhD thesis, Harvard University. 12

Williams, R. J. and Zipser, D. (1995). *Gradient-Based Learning Algorithms for Recurrent Networks and Their Computational Complexity*, page 433–486. L. Erlbaum Associates Inc. 19

Wolanin, A., Mateo-Garciá, G., Camps-Valls, G., Gómez-Chova, L., Meroni, M., Duveiller, G., Liangzhi, Y., and Guanter, L. (2020). Estimating and understanding crop yields with explainable deep learning in the Indian Wheat Belt. *Environmental Research Letters*, 15(2). xix, 97, 113

Wu, B., Meng, J., Li, Q., Yan, N., Du, X., and Zhang, M. (2014). Remote sensing-based global crop monitoring : experiences with china's cropwatch system. *International Journal of Digital Earth*, 7(2) :113–137. xvii

Xu, K., Ba, J., Kiros, R., Cho, K., Courville, A., Salakhudinov, R., Zemel, R., and Bengio, Y. (2015). Show, attend and tell : Neural image caption generation with visual attention. In *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pages 2048–2057, Lille, France. PMLR. 25

You, J., Li, X., Low, M., Lobell, D., and Ermon, S. (2017). Deep Gaussian process for crop yield prediction based on remote sensing data. *31st AAAI Conference on Artificial Intelligence, AAAI 2017*, pages 4559–4565. xix, 97, 113

Yuan, Y. and Lin, L. (2021). Self-supervised pretraining of transformers for satellite image time series classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14 :474–487. 123

Zhang, L., Song, J., Gao, A., Chen, J., Bao, C., and Ma, K. (2019). Be your own teacher : Improve the performance of convolutional neural networks via self distillation. In *2019 IEEE/CVF International Conference on Computer Vision, ICCV 2019, Seoul, Korea (South), October 27 - November 2, 2019*, pages 3712–3721. IEEE. 72

Zhou, Y. and Chellappa, R. (1988). Computation of optical flow using a neural network. *IEEE 1988 International Conference on Neural Networks*, pages 71–78 vol.2. 17

Acronymes

AA Apprentissage automatique.

AP Apprentissage profond.

AVHRR Advanced Very-High-Resolution Radiometer.

CESBIO Centre d'Etudes Spatiales de la Biosphère.

Cirad Centre de Coopération Internationale pour la Recherche Agronomique et le Développement.

CNES Centre National d'Études Spatiales.

CNN Convolutional Neural Network ou Réseau de neurones convolutionnels.

DL Deep learning.

ESA European Spatial Agency.

FAO Organisation des Nations Unies pour l'Alimentation et l'Agriculture.

GMES Global Monitoring for Environment and Security.

GPS Global Positioning System.

GRD Ground Range Detected.

GRU Gated Recurrent Unit.

HH Polarisation horizontale pour la transmission et la réception.

HV Polarisation horizontale pour la transmission et verticale pour la réception.

IA Intelligence artificielle.

IDS Infrastructure de Données et de Services.

IRSA Institute of Remote Sensing Applications of the Chinese Academy of Sciences.

IW Interferometric Wide swath.

JRC Joint Research Centre.

LCCS Land Cover Classification System.

LSTM Long-Short Term Memory.

MA Mécanisme d'attention.

MLP Multi-layer Perceptron ou Perceptron multi-couche.

MODIS Moderate-Resolution Imaging Spectroradiometer.

MSI MultiSpectral Instrument.

NOAA National Oceanic and Atmospheric Administration.

OBIA Object Based Image Analysis.

ONU Organisation des Nations unies.

OT Observation de la Terre.

PEPS Plateforme d'Exploitation des Produits Sentinel.

ReLU Rectifier Linear Unit.

RNN Recurrent Neural Network ou Réseau de neurones récurrents.

RPG Régistre Parcelle Graphique.

RSO Radar à synthèse d'ouverture.

SAR Synthetic-Aperture Radar.

SPOT Système probatoire d'observation de la Terre ou Satellite pour l'observation de la Terre.

THEIA Pôle de données et de services surfaces continentale.

TOA Top of Atmosphere.

TOC Top of Canopy.

UNESCO Organisation des Nations unies pour l'éducation, la science et la culture.

USAID Agence des États-Unis pour le développement international.

VH Polarisation verticale pour la transmission et horizontale pour la réception.

VV Polarisation verticale pour la transmission et la réception.