Mémoire présenté pour obtenir le diplôme de MASTER 2 Mention GAED

Spécialité TELENVI (Télédétection/Environnement)

Université Rennes 2 et Institut Agro Rennes/Angers

Titre :

Estimation du yield gap et cartographie des niveaux d'intensification par télédétection et méthodes de machine learning

Présenté et soutenu par :

M. Mansour DIOP

Le 26 Septembre 2024

Devant le jury composé de :

M.FOUAD Youssef	Enseignant-Chercheur – Institut Agro	Président du jury
M. Jérémy LAVARENNE	Chercheur - CIRAD	Encadreur
Mme. L.Hubert MOY	Enseignant-Chercheur – Univ Renne2	Membre
Mme. C. LARGOUET	Enseignant-Chercheur – Institut Agro	Membre
Mme.C.BISSUEL	Enseignant-Chercheur – Institut Agro	Membre



DEDICACES ET REMERCIMENTS

Au terme de cette étude, je rends d'abord grâce à Allah le tout puissant de m'avoir donné la santé et l'aptitude intellectuelle d'accomplir ce travail.

Je remercie aussi à tout ce qui ont contribué à ma formation et à cette étude :

M. Youssef Fouad, Responsable de Master TELENVI

Mme. Laurence Hubert Moy, Professeur TELENVI et

M. Jérémy LAVARENNE, encadrant et chercheur au CIRAD

M. Simon MADEC, Chercheur au CIRAD

Je dédie ce travail à mon guide spirituelle **Mawlaya Cheikh Seyidi Mouhamadoul Moustapha SY Al MAKTOUM, à mon Capitaine Mame Cheikh Ahmed Tidiane SY Al Moustapha** que Allah vous accorde la longévité et la santé dans l'accomplissement de tous vos projets.

A ma très chère mère chérie **Mbene Tall et** à mon père **Mor Talla DIOP** qui se battent corps et âme pour la réussite de leurs enfants, que le bon Dieu Allah vous accorde la longévité, la santé et que vous soyez présents pour assister notre réussite dans les meilleures situations.

A mes sœur set mes frères : Mamour DIOP, Pape saliou DIOP, Yacine DIOP, Mouhamed Lamine DIOP et Yacine seye DIOP vous êtes ma raison de vivre.

A mes condisciples et frères spirituels *:* Abdou MBAYE, Elhadj Assane MBOUP, Sokhna Binta, Sokhna Oumou, Sokhna Astou, Sokhna Mariama et tout le **Dahira Moustarchidine wal Moustarchidate.**

A mes meilleur et très amis Cheikh Tidiane Ndao, Daouda Seck et Serigne Babacar Seck qui sont à la fois mes frères spirituels et temporels.

RESUME

Etant une zone où la sécurité alimentaire demeure un enjeu crucial, l'Afrique subsaharienne est caractérisée par des cultures de subsistance (agriculture familiale) qui varient d'une région à l'autre. Ces systèmes agricoles sont parfois très complexes à étudier en raison de plusieurs facteurs endogènes très hétérogènes. Ces facteurs peuvent être de natures agronomiques et/ou socioéconomiques et impactent directement sur le rendement des cultures. Cependant, Plusieurs systèmes d'alerte précoce (SAP) ont été développés pour anticiper les crises et planifier des mesures d'urgence. Toutefois, ces SAP présentent des alertes contrastées et parfois contradictoires, et donc une alternative est d'étudier les yield gap et les niveau d'intensification. Ces derniers peuvent être considérés comme des paramètres représentant de manière synthétique des facteurs agronomiques et socioéconomiques expliquant l'écart entre le rendement atteignable et rendement réel. Dans cette étude nous tirons parti d'un jeu de données de rendements déclarés spatialisé, et de données issues de télédétection au travers de méthodes de machine learning (ML) pour évaluer les possibilités de réalisation de cartographie du yield gap de niveau II+III à l'échelle de la bande soudano-sahelienne, une sous-région particulièrement concernée par les problématiques de sécurité alimentaire. Des indices de végétation et des paramètres du sol issus de la télédétection sont utilisés avec des modèle ML comme le Random forest (Rf), le Xgboos le ADaBoost (ADB) et le Decision Tree (DT). Nos résultats montrent que l'approche RF retourne des performances assez correctes dans l'estimation du yield gap avec un R^2 = 76,54% et avec des valeurs très faible de moins de -500 Kg/ha traduisant des niveaux d'intensification élevé localisés par exemple au nord du Bénin et le long du fleuve Niger.

ABSTRACT

As an area where food security remains a crucial issue, sub-Saharan Africa is characterised by subsistence farming (family farming) that varies from one region to another. These farming systems are sometimes very complex to study due to a number of highly heterogeneous endogenous factors. These factors may be agronomic and/or socio-economic in nature and have a direct impact on crop yields. However, several early warning systems (EWS) have been developed to anticipate crises and plan emergency measures. However, these EWS present contrasting and sometimes contradictory alerts, so an alternative is to study yield gaps and intensification levels. The latter can be considered as parameters that synthetically represent the agronomic and socio-economic factors that explain the gap between achievable yield and actual yield. In this study, we take advantage of a spatially declared yield dataset and remote sensing data using machine learning (ML) methods to assess the possibilities of mapping the level II+III yield gap at the scale of the Sudano-Sahelian band, a sub-region particularly concerned by food security issues. Vegetation indices and soil parameters derived from remote sensing are used with ML models such as Random Forest (Rf), Xgboos, ADaBoost (ADB) and Decision Tree (DT). Our results show that the RF approach performs fairly well in estimating the yield gap, with an R2 = 76.54%, and with very low values of less than -500 kg/ha, reflecting high levels of intensification located, for example, in northern Benin and along the Niger River.

TABLE DES MATIERES

LISTE DES FIGURES

LISTE DES ACRONYMES

SAP : Système d'Alerte Précoce

NDVI : Normlized Différence Vegetation Index

SAVI : Soil Adjusted Vegetation Index

MSAVI : Modified Soil Adjusted Vegetation Index

OSAVI : Optimized Soil Adjusted Vegetation Index

ARVI : Atmosphere Resistant Vegetation Index

EVI : Enhanced Vegetation Index

LAI : Leaf Area Index

ML : Machine Learning

RMSE : Root Mean Square Error

MAE : Mean Absolute Error

RF : Random Forest

XGB : XGBoost

ADB : ADaBoost

DT : Decision Tree

INTRODUCTION GENERALE

La sécurité alimentaire en Afrique subsaharienne représente un enjeu critique, influencé par une multitude de facteurs complexes et interconnectés. La région fait face à des défis majeurs tels que la croissance démographique rapide, une agriculture de subsistance dépendante des précipitations, ainsi que des risques sanitaires et sécuritaires. Cette insécurité alimentaire s'aggrave tout particulièrement en Afrique subsaharienne, où la proportion de la population souffrant de faim chronique est la plus élevée au monde (Vintrou, 2013). Dans ce contexte, la capacité à anticiper les crises alimentaires et à planifier des interventions d'urgence est cruciale, et il existe un besoin de suivi systématique et précis des systèmes agricoles et de leur adaptation à un environnement en évolution, pour pouvoir évaluer les impacts sur la sécurité alimentaire (Vintrou, 2013). Pour ce faire, divers systèmes d'alerte précoce (SAP) ont été développés. Ces système d'alerte précoce sont des outils de rationalisation de l'attribution de l'aide alimentaire (François Enten, 2008) visant à fournir une information fiable et préventive sur les risques potentiels de crise alimentaire ou d'insécurité alimentaire localisée (Brown et al., 2007). Cependant, ces systèmes présentent souvent des alertes contrastées et parfois contradictoires, rendant difficile une réponse unifiée et efficace aux crises alimentaires. Parmi ces SAP, le FEWS NET (Famine Early Warning Systems NETwork) de l'USAID (United States Agency for International Development) ou le GIEWS (Global Information and Early Warning System) de l'ESA (Agence Spatiale Européenne) utilisent des méthodes différentes et complémentaires pour estimer des rendements (Vintrou, 2013). Cependant, les modèles de simulation de culture peuvent apporter un point de vue complémentaire en estimant le rendement atteignable des cultures (c'est-à-dire le rendement que les agriculteurs peuvent raisonnablement espérer atteindre compte tenu des pratiques agricoles locales, des variétés cultivées, et des conditions climatiques spécifiques d'une saison) en fonction de données agropédoclimatiques, et de paramètres variétaux. Parmi ces modèles, SARRA-H est utilisé en Afrique de l'Ouest depuis plus de 30 ans (Baron, 2013). Ces rendements simulés, confrontés aux rendements réels (observés), permettent l'estimation des écarts de rendement (yield gap) des cultures, fournissant ainsi une base pour des stratégies d'intervention plus précises. Une approche prometteuse consiste à intégrer le concept de "niveau d'intensification" des cultures. Ce concept est considéré comme l'absence de toute contrainte hydrique du fait de la fertilité du sol et des techniques de gestion utilisées (Affholder, 1997). Parallèlement ces facteurs agronomiques et socioéconomiques sont les causes principales des écarts de rendement (yield gap) et donc par

conséquent il existe une relation étroite entre ce yield gap et le niveau d'intensification des cultures.

Dans le cadre de cette étude, les données issues de la télédétection satellitaire peuvent jouer un rôle essentiel en fournissant des informations précises sur la densité, la conduite, et la dynamique du couvert végétal pour analyser le yield gap. Ces données permettent de formuler des hypothèses de départ robustes pour étudier les niveaux d'intensification des culture. L'hypothèse globale de cette étude est : les données de télédétection satellitaire peuvent être utilisées pour estimer le niveau d'intensification des cultures.

Le présent projet de recherche vise à estimer le yield gap et à cartographier un indice de niveau d'intensification des cultures dans les systèmes agricoles d'Afrique de l'Ouest en utilisant des données de télédétection. Pour ce faire, des méthodes d'apprentissage automatique seront employées pour construire des modèles à partir des séries temporelles MODIS, des paramètres physico-chimique du sol spatialisés , ainsi que des données issues d'enquêtes.

Problématique et Objectifs de la Recherche

L'un des principaux défis de la sécurité alimentaire en Afrique subsaharienne est la variabilité des rendements agricoles. Cette variabilité est exacerbée par des pratiques agricoles hétérogènes et souvent non intensifiées. L'objectif principal de cette recherche est de développer des modèles d'estimation du yield gap et des modèles pour cartographier le niveau d'intensification des cultures à partir de données de télédétection, afin de fournir des cartes détaillées qui pourront être utilisées pour améliorer la planification agricole et les interventions d'urgences.

Les objectifs spécifiques de ce projet sont les suivants:

- Calculer des yield gap
- Utiliser des méthodes d'apprentissage automatique pour entraîner des modèles de spatialisation des yield gap et d'estimation du niveau d'intensification des cultures.
- Évaluer et valider les modèles de régression et de classification obtenus.
- Produire des cartes de yield gap et de niveau d'intensification basées sur les données disponibles.

CHAPITRE I : ETAT DE L'ART

En Afrique Sub-Saharienne, plusieurs pays et en particulier ceux du Sahel sont confrontés à un déséquilibre alimentaire lié à un déficit régulier de leur production agricole (Traore, 2022). Ce déficit alimentaire régulier ne permet pas d'assurer la sécurité alimentaire. Elle se définis comme l'existence pour tous les êtres humains, à tout moment, d'un accès physique et économique à une nourriture suffisante, saine et nutritive qui répond à leurs besoins et préférences alimentaires pour mener une vie saine et active (FAO, 2006). La sécurité alimentaire est donc devenue un défi majeur pour ces pays qui doivent produire davantage pour nourrir une population en forte croissance (FAO, 2016a). Les agriculteurs soudano-sahéliens en particulier ceux d'Afrique de l'Ouest cultivent principalement du mil, du sorgho et du maïs comme cultures de base avec un minimum d'intrants et d'équipements, ce qui se traduit le plus souvent par un faible rendement (Traore, 2022). Ces cultures à l'instar du mil et du sorgho présentent de nombreux avantages en termes d'adaptation aux conditions climatiques extrêmes et est caractérisée par une forte capacité de résilience au déficit hydrique (Tadele, 2017).

L'amélioration de cette sécurité alimentaire en Afrique de l'ouest nécessitera une bonne compréhension des rendements potentiels des systèmes agricoles ainsi que la réduction des écarts de rendement (c'est-à-dire la différence entre le rendement potentiel et le rendement réel des agriculteurs)(Khechba *et al.*, 2021). A cette fin, l'étude des écart de rendement dans les pays africains devra mettre l'accent sur l'impact des pratiques culturales ainsi que les aspects socioéconomiques pour combler cet écart. Dans la plupart des pays africain, les systèmes de production sont dominés par des petits exploitants qui observent des écarts importants entre le rendement atteignable et le rendement observé. Il est donc important d'étudier les facteurs de ces écarts de rendement. Il est à noter que dans ces pays où les données agricoles sont souvent de qualité faible, les études d'analyse des écarts de rendement peuvent être inexactes, en particulier là où les terrains agricoles présentent une grande variabilité et complexité en termes de terres cultivées et de propriétés du sol(Khechba *et al.*, 2021). C'est dans ce sens que les projets LSMS-ISA et RHOMIS de la banque mondiale viennent en appuis en fournissant des données de précision dans le temps et dans l'espace, et représentent une source de donnée exhaustive pour les rendements observés. Cependant, spatialiser ces données matérialisées par des points de rendement en kg/ha issues de ces projets d'enquêtes pour mieux appréhender la variabilité de ce rendement requière une mobilisation d'outil comme la télédétection et des méthodes de modélisation.

Compte tenu de la nécessité de méthodes utiles pour aider à identifier les régions ayant le plus grand potentiel d'augmentation de l'approvisionnement alimentaire en Afrique grâce à la minimisation des écarts de rendement, les techniques de télédétection se sont révélées particulièrement utiles pour surveiller et analyser les rendements des cultures au cours des dernières décennies, en raison de leur capacité à traiter les données spatiales à grande échelle et à fournir des résultats qui peuvent être modélisés (Dehkordi *et al.*, 2020). De plus une évaluation quantitative et une cartographie des propriétés clés du sol et des cultures telles que le NPK à l'aide de données de télédétection fournie des informations intéressantes pour réduire l'écart de rendement (Lobell, Cassman and Field, 2009a). Ces informations spatialisées détaillées sur les propriétés physico-chimique du sol sont essentielles pour mener une analyse des écarts de rendement surtout en Afrique où la dégradation des terres et la perte de fertilité des sols ont été signalées par de nombreuses études (*weber, et al, 2018*). Plusieurs méthodologies ont été développées pour estimer l'écart de rendement des cultures, en utilisant des approches telles que les enquêtes, l'expérimentation sur le terrain, la télédétection, l'apprentissage automatique ou une combinaison de ces méthodes. L'apprentissage automatique, tel que le Random forest, est utile pour comprendre et classer l'importance relative des facteurs contribuant aux rendements des cultures (Amgain *et al.*, 2021). L'intégration de la télédétection avec l'apprentissage automatique se distingue par sa capacité à prédire les rendements avec précision en analysant le NDVI (indice de végétation par différence normalisée) en cours de saison et d'autres valeurs de l'indice de végétation. Cette approche, exploitant particulièrement les indices de végétation issus des données de télédétection, a été largement appliquée à diverses cultures, notamment le maïs (Li *et al.*, 2022), (Shuai and Basso, 2022). En plus d'avoir une forte capacité à estimer les rendements observés, la télédétection constitue un outil efficace pour évaluer les rendements potentiels des cultures(Wang et al., 2023), (Lobell, 2013), en utilisant souvent le percentile élevé (par exemple, 95e) des distributions de rendement pour déterminer l'écart de rendement au lieu de s'appuyer uniquement sur les rendements les plus élevés enregistrés dans une zone donnée. Cependant, des modèles de simulation basés sur des processus tels que Sarra-H sont utilisés pour simuler des rendements atteignables. (Sultan, Defrance and Iizumi, 2019) ont implémenté le modèle Sarra-H pour simuler le rendement des cultures des zones tropicales sèches, telles que le mil et le sorgho et ont évalué l'impact du changement climatique sur les pertes de rendement. Leur étude a révélé des pertes de 17,7% pour le mil et de 15% pour le sorgho, et les pertes les plus importants sont localisées au nord du sahel.

Il est à noter que ces modèles de simulation basés sur les processus sont généralement conçu pour évaluer le potentiel des cultures régionales en fonction des caractéristiques du sol et du climat, et des principales caractéristiques des cultivars. Ces modèles de simulation de culture peuvent apporter un point de vue complémentaire en estimant le rendement atteignable des cultures. Et donc passer du rendement atteignable à une estimation du rendement réel peut alors être de considérer l'utilisation du concept de "niveau d'intensification", qui peut être considéré comme un paramètre représentant de manière synthétique des facteurs agronomiques et socioéconomiques. En outre ce niveau d'intensification est perçu comme le rendement potentiel qui serait obtenu sur un champ en l'absence de toute contrainte hydrique du fait de la fertilité du sol et des techniques de gestion utilisées (Affholder, 1997). Dans le cadre de cette étude, ce niveau d'intensification est le nombre de classe du yield gap issu de sa discrétisation. Et donc il sera question d'explorer les indices végétaux et les paramètres du sol pour voir s'ils possèdent des informations pertinentes pour expliquer ce niveau d'intensification des cultures dans la bande sahélienne.

CHAPITRE II : MATERIELS ET METHODES

En vue de cartographier le niveau d'intensification des cultures dans la bande sahélienne, le calcul et la spatialisation du yield gap est nécessaire. Il sera donc primordiale de procéder à la simulation de rendement atteignable (Ra) par modélisation avec le model sarra-py et utiliser une approche delta (Ra – Rd) pour le calcul du yield gap. Ainsi, des produits de télédétection tels que des indices de végétation et des paramètres physico-chimiques du sols sont intégrés dans des méthodes de machines learning (Random forest, Decision tree, XGBoost…) pour créer des modèles de yield gap et de niveau d'intensification. Cette démarche méthodologique permettra de produire des cartes de yield gap, de niveau d'intensification et d'identifier quelles sont les covariables qui influencent le plus sur ces niveaux d'intensification.

II.1 Zone d'étude

Cette étude est portée sur la bande sahélienne, située dans l'emprise géographique 15.7, -12.5, 8.3 et 14.8 (latitude Nord, longitude nord, latitude sud et longitude sud), souvent qualifiée de zone de transition entre le désert et les prairies du Sahel qui abrite un climat semi-aride avec des vents chauds et secs (Wu *et al.*, 2020). Le relief de cette zone comprend des plateaux ondulés. Les précipitations varient considérablement du nord au sud, de 150mm à 600mm/an, avec dans le même temps des précipitation variant considérablement entre la saison sèche et la saison humide ; principalement concentrées entre juin et septembre (Wu *et al.*, 2020).

L'emprise géographique de l'étude couvrant plusieurs pays, néanmoins notre travail cible principalement quatre pays que sont : la Mali, le Burkina Faso, le Niger et le Bénin. La figure1 ci-dessous montre la cartographie de la zone d'étude matérialisée par un contour rouge, et la représentation de la topographie de la zone d'étude.

Figure 1 : Localisation géographique de la zone d'étude

II.2. Données de rendement déclaré: base de données LSMS/ISA et RHoMIS

Dans cette étude, un jeu de donnée harmonisé sur les rendements du sorgho et du mil en Afrique de l'Ouest (doi:10.5281/zenodo.10556265) intégrant les enquêtes LSMS-ISA et RHoMIS est utilisé en guise de rendement déclaratif. Ces données comprennent des déclarations de production et de surfaces emblavées permettant de calculer des rendements, mais aussi des informations sur la production agricole, les pratiques de fertilisation et les coordonnées GPS des parcelles. La Figure 2 représente les localisations auxquelles nous disposons d'au moins une valeur de rendement déclaré.

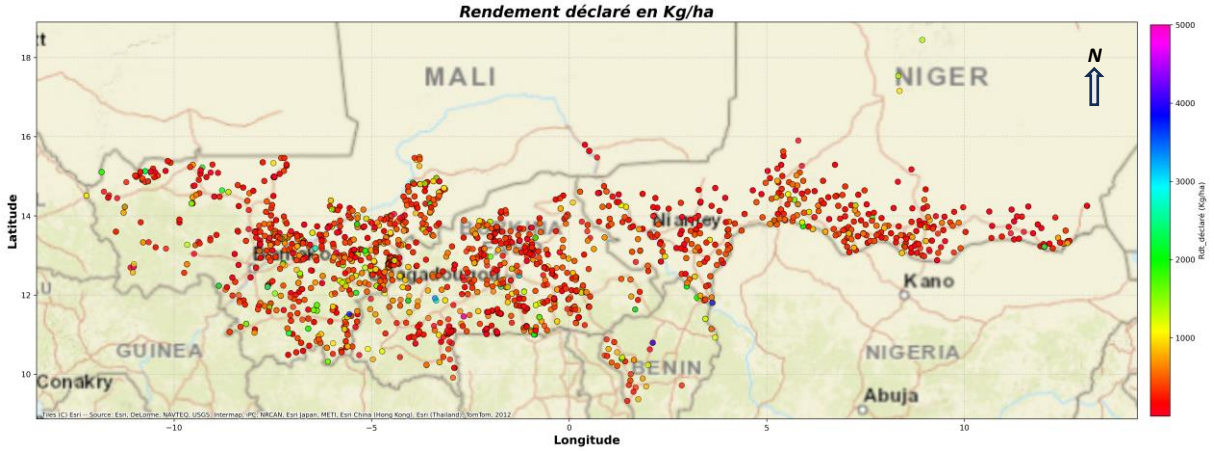


Figure 2: Répartition spatiale des données de rendement observé

Dans la suite de l'étude, le rendement du mil est retenu car étant plus représentatif dans l'espace que celui du sorgho et couvrant la plus grande emprise spatiale, soit 9543 points d'observation après nettoyage. En observant ces données de rendement du mil, et la Figure 3, il s'avère que plus de 88% du rendement du mil est inférieur à 1000 Kg/ha. Par ailleurs, les valeurs maximales peuvent atteindre jusqu'à 5000 Kg/ha.

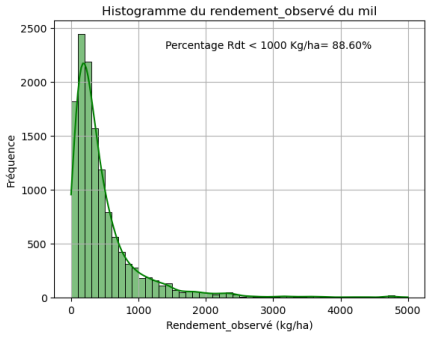


Figure 3: Histogramme des données de rendements déclarés du mil

II.3 Simulation des rendements potentiels avec SARRA-Py

SARRA-Py est un modèle de simulation de culture implémenté en langage Python à partir de la famille de modèle SARRA (SARRA-H, SARRA-O) (Baron Cristian, 2013). Il se base sur un modèle de bilan hydrique dynamique à pas de de temps quotidien, utilisé pour estimer l'impact des scénarios climatiques sur les cultures annuelles, en supposant que la performance des cultures est fonction des contraintes hydrologiques accumulées au cours du cycle de croissance de la culture. Dans le cadre de cette étude, le modèle SARRA-Py version 0.1.1 (doi:10.5281/zenodo.10969997) a été utilisé pour simuler le rendement atteignable dans toute la bande sahélienne. Pour ce faire, les entrées du modèle que sont les données climatiques et météorologiques sont téléchargées à partir du 1er avril 2018 jusqu'au 31 décembre 2018. Cela est facilité par l'outil SARRA-data-download qui, utilisé en v0.6.2 (doi:10.5281/zenodo.11365141) a permis de télécharger pour toute l'emprise spatiale de la zone d'étude des estimations quotidiennes des précipitations TAMSAT, ainsi que les variables météorologiques nécessaires au calcul de l'évapotranspiration de référence ET0 par la formule de Hargraeves (calcul effectué dans SARRA-data-download) depuis Copernicus AgERA5 (doi: 10.24381/cds.6c68c9bb).

Cette étape est suivie de la simulation du rendement atteignable, cependant en utilisant deux méthodes visant à explorer l'impact de deux sources de données décrivant la capacité de rétention en eau du sol. Ces deux méthodes intègrent des fonctions qui visent à charger une carte des propriétés physiques du sol qui sont utilisées pour obtenir une capacité de rétention d'eau du sol. Cette capacité de rétention d'eau est utilisée dans les calculs du bilan hydrique quotidien du sol. La première méthode ('méthode alternative ') utilise des cartes de propriétés du sol pour le calcul de la réserve utile. En revanche, la deuxième méthode (méthode avec Ru_ISRIC) va charger la réserve utile à partir d'une couche qui est fournie telle que calculée par l'ISRIC. Globalement, cela signifie que la capacité de rétention d'eau du sol de la seconde méthode est fournie par un produit cartographié et sera plus continue dans l'espace, alors que la première méthode entraînera des ruptures dans les cartes, d'où quelques artefacts.

II.4. Choix des covariables utilisés pour spatialiser le yield gap et modéliser le niveau d'intensification des cultures

Les covariables qui sont utilisés dans cette études sont des indices de végétation et des paramètres du sol calculés depuis les API ***Geemap et ISDA_SOIL*** implémentés en python. Pour les indices de végétation, tenant compte de l'échelle de cette étude (échelle régionale) notre choix est porté sur les produits dérivés du satellite **MODIS**. Ce dernier collecte en

permanence des données dans 36 canaux spectraux avec une couverture mondiale tous les 1 à 2 jours et une résolution spatial de 500m. Sa gamme spectrale large permet aux données d'être utilisées dans des études de nombreuses disciplines, notamment la santé végétative, les changements dans la couverture terrestre et la prédiction des rendement (Petersen, 2018).

Les indices de végétation peuvent constituer un outil précieux car elles sont utiles à la prise de décision, à la gestion des cultures, à la planification des récoltes, à la prévision du rendement des cultures, à la collecte et au suivi des informations (Barboza *et al.*, 2023). Dans cette étude, les indices (NDVI, LAI, EVI, SAVI, MSAVI, OSAVI, ARVI) sont utilisé en raison de leur lien statistique à certaines caractéristiques du couvert végétal telles que la surface foliaire et la biomasse végétative. Ainsi, la moyenne de ces indices est calculée dans la période végétative mai – octobre 2018.

En effet, l'indice NDVI peut quantifier la verdure de la végétation, comprendre la densité de la végétation et évaluer les changements dans la santé des plantes (Zsebő *et al.*, 2024). Plusieurs études ont déclaré qu'une corrélation significative ($p \leq 0,05$) pouvait être trouvée entre l'indice NDVI et le niveau d'azote (Teboh *et al.*, 2011) (Kizilgeci *et al.*, 2021).

D’autres indices, dérivés directs ou indirects de l’information spectrale ou des indices de végétation, sont utilisés dans cette étude. Notamment, des estimations de LAI qui est un paramètre biophysique renseignant sur la couverture du feuillage et la croissance des culture. (Parra, Borrás and Gambin, 2022) ont conclu dans leur travaux que l’étude du Lai est important et pourraient contribuer à augmenter les rendements du sorgho tout en anticipant l’anthèse. Et par conséquent, nous supposons que les valeurs du LAI peuvent être corrélées à certaine zone où il y’aurait un retard ou des stades de floraison assurés de façon optimale par les cultures et donc traduisant un niveau d’intensification élevé.

Des couches d’informations indirectement issues de télédétection sont également utilisées : Des cartes de propriétés physico-chimiques des sols (texture (% de silt et %clay), le pH, le carbone organique (CO), l’azote (N) fournissent une évaluation quantitative et une cartographie de ces propriétés clés du sol s'avèrent très importante pour l’étude des écarts de rendement (Lobell, Cassman and Field, 2009b).

Le carbone organique du sol est une clé de la fertilité. En effet, il subit dans le sol des transformations biologiques qui mènent à leur minéralisation et à la libération des éléments minéraux tels que l’azote, le phosphore, le soufre, le potassium, et des oligoéléments, qui

deviennent disponibles pour les plantes. Son abondance dans le sol peut être favoriser par des amendements organiques ;

Le pourcentage de la fraction argileuse du sol est aussi un paramètre physique du sol capable de protéger l'humus de l'action des microorganismes en ralentissant le processus de minéralisation pour la formation du complexe argilo-humique.

Le pH, un paramètre chimique du sol essentiel, détermine la qualité de vie des microorganisme du sol pour une bonne assimilation des éléments minéraux par les plantes.

Donc il est pertinent d'intégrer certains paramètres du sol dans la modélisation des écarts de rendement et du niveau d'intensification. Ces paramètres sont récupérés depuis l'API ISDA_SOIL qui fournit des produit avec 30m de résolution. Ces paramètres sont rééchantillonné par la suite à la résolution des produits de modis (500m) pour la création des stacks.

Tableau 1 : Covariables utilisés pour la spatialisation du yield gap et la modélisation du niveau d'intensification des cultures

Type de covariable	Covariable	Organisme producteur	Resolution spatio-temporelle	Link/DOI	Formules
Indices de végétation	NDVI	NASA (MODIS)	500m, 2j	https://doi.org/10.5067/MODIS/MOD13A1.061	$NDVI = \frac{(PIR - ROUGE)}{(PIR + ROUGE)}$
	LAI	NASA(MODIS)	500m, 2j	https://doi.org/10.5067/MODIS/MCD15A3H.061	
	EVI	NASA(MODIS)		https://doi.org/10.5067/MODIS/MOD13A1.061	$EVI = 2,5\frac{(NIR - Rouge)}{(1 + NIR + 6Rouge - 7,5Bleu)}$
	SAVI	NASA(MODIS)	500m, 2j	https://doi.org/10.5067/MODIS/MOD13A1.061	$SAVI = \frac{(NIR - R)}{(NIR + R + L)} \times (1 + L)$
	MSAVI	NASA(MODIS)	500m, 2j	https://doi.org/10.5067/MODIS/MOD13A1.061	$MSAVI = \frac{2 * NIR + 1 - \sqrt{(2 * NIR + 12 - 8 * (NIR - rouge)}}{2}$
	OSAVI	NASA(MODIS)	500m, 2j	https://doi.org/10.5067/MODIS/MOD13A1.061	$OSAVI = \frac{NIR - Rouge}{(NIR + Rouge + L)} * 1 + L$
	Arvi	NASA(MODIS)	500m, 2j	https://doi.org/10.5067/MODIS/MOD13A1.061	$ARVI = \frac{(NIR - (2Rouge - Bleu))}{(NIR + (2Rouge - Bleu))}$

Propriétés de sol	CO, pH, N, %Clay, %Silt	ISDA_SOIL	30m	https://www.isda-africa.com/isdasoil/	

II.5. Détermination du yield gap et discrétisation

Après avoir simuler le rendement atteignable (Ra) par le model sarra-py, une extraction des valeurs est faite au coordonnées de chaque point des rendements déclaré (Rd). Cette tâche est réalisée à l'aide de l'outil *point sampling tools* de Qgis version 3.28. Le yield gap est calculé en faisant la différence (Ra – Rd). Pour matérialiser les classes de niveau d'intensification, Ce yield gap est ensuite discrétisé en 4 classes (élevée, moyenne, faible et très faible) par la méthode quantile via la fonction qcut() de la librairie Pandas. La classe 'élevée' représente la classe de niveau d'intensification plus élevée, et correspond aux yield très faibles. Ce qui signifie donc que les valeurs de rendements déclaré atteignent ou dépassent les valeurs de rendements simulées.

II.6.Extraction et préparation des covariables utilisés pour l'entrainement des modèles

Une fois avoir calculé pour chaque localisation un yield gap et un niveau d'intensification correspondant, une extraction des valeurs des covariables est réalisée. Cependant, dans la mesure où les coordonnées des données de déclaration de ménage sont celles de l'habitation, une grande partie des points n'est pas géolocalisés sur des zones de culture mais plutôt sur des zones artificialisées (bâti). De fait, et afin d'extraire des valeurs de covariables qui ne soient pas influencées par les zones urbaines, nous appliquons la stratégie d'extraction suivante : nous allons supposer que sur un rayon de 5 km du point, que l'ensemble des cultures prennent pour valeur de rendement celle déclarée en ce point. En outre, une masque des zones de cultures dérivée du produit esa_landcover_2018 (https://zenodo.org/records/3518038) est appliqué pour ne retenir que les pixels des covariables se trouvant effectivement superposés à une zone cultivée supposée. Ainsi, la valeur des covariables est extraite à condition que le buffer intercepte la couche de la zone de culture, ce qui permet de mettre en correspondance la moyenne des rendements déclarés en ce point depuis les données LSMS-ISA avec une zone où une activité agricole est plausible, et où le risque que les covariables issues de la télédétection renvoient un signal urbain est limité.

II.7. Entrainement et validation de modèles de ML

Après l'extraction des valeurs des covariables, deux jeux de données sont construits pour la modélisation. Ces jeux de données sont tels que: un avec comme variable cible le yield gap et l'autre avec comme variable cible les classes de niveau d'intensification issu de la discrétisation.

Avant l'étape d'entrainement des modèles, les jeux de données sont séparés en jeux d'entrainement et de validation avec l'option stratification avec la fonction *stratifiedGroupKfold* dans *scikt-learn (version=1.5.1)* en 80% de jeux de calibration et 20% de jeux de validation. Cette méthode de stratification tente de créer des plis stratifiés avec des groupes non supervisés dans le but d'éviter des jeux de train et de test trop proches géographiquement. Il s'en suit une étape de normalisation des valeurs des variables explicatives avec la méthode de *standarscaler* dans *sklearn* qui a comme fonction de centrer et réduire les données. En effet, cette méthode de normalisation est essentielle dans cette étude car les variables explicatives pour le niveau d'intensification des cultures sont de caractéristiques et d'unités différentes.

Quatre types de méthodes de machine learning à savoir, le Random forest, le decision tree, le XGboost et le ADaboost ont été implémentées pour tenter de prédire le yield gap (tâche de régression) et le niveau d'intensification (tâche de classification). Tous les algorithmes utilisés sont tels qu'implémentés dans la librairie python *scikt-learn (version=1.5.1).* Concernant la stratégie d'entraînement de multiples modèles, à l'entame du processus de modélisation, un modèle dit "baseline" de très faible complexité est implémenté, comprenant donc peu de variables explicatives et utilisant des hyperparamètres par défaut pour l'entraînement. Dans un second temps, d'autres modèles plus complexes sont entraînés avec l'ajout de covariables pertinentes, et la fonction *randomizedSearchCV* est utilisée afin de trouver les hyperparamètres optimaux dans le but d'obtenir les meilleurs performances.

La stratégie de la cross validation a été adoptée dans cette étude pour évaluer la précision de tous les modèles ML. L'ensemble du jeu de donnée est divisé en 10 sous-ensemble. Un sous ensemble pour la validation et neuf pour la calibration. Cela est répété 10 fois, en utilisant à chaque fois un ensemble de données de calibration et de validation différent. Le résultat de l'évaluation des performances d'un k-fold est la moyenne des performances obtenues sur les 10 ensembles d'entrainement et de test.

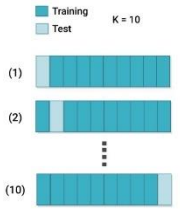


Figure 4 : schéma de principe de la cross validation. Dans cet exemple à chaque fold, 90% des données sont sélectionnées pour l'entraînement et 10°% écartées pour l'évaluation des performances. Au fold suivant on réitère l'opération en excluant les valeur

II.7.1.Validation des méthodes de régression

Les performances des méthodes de régression ont été évaluées en terme de coefficient de détermination (R^2), de la racine carré de l'erreur quadratique moyenne (RMSE en kg/ha) et de l'erreur absolue moyenne (MAE en kg/ha) :

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(yi - \hat{y}i)^2}{\sum_{i=1}^{n}(yi - \bar{y}i)^2}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(yi - \hat{y}i)^2}{n}}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|yi - \hat{y}i|$$

Avec yi la valeur observée du yield gap, $\hat{y}$i la valeur du yield gap prédite, $\bar{y}$i sa valeur moyenne observée et n le nombre d'observations.

II.7.2.Validation des méthodes de classification

Pour valider les modèles de classifications, des métriques tels que l'accuracy, le F1 score et le coeficcient Kappa sont calculé.

L'accuracy est le rapport entre les prédictions correctes et toutes les prédictions faites par un algorithme. Le score F1 combine les informations de précision en une seule mesure allant de 0 à 1 et prend en compte à la fois la précision et le rappel. Il est la moyenne harmonique de la précision et du rappel. Il s'agit d'une bonne métrique équilibrée de faux positifs et de faux négatifs.

Le coefficient Kappa un score qui exprime le niveau d'accord entre deux annotateurs sur un problème de classification. Il est défini comme :

$$\boldsymbol{k = (po - pe) / (1 - pe)}$$

Ou ***po*** est la probabilité empirique d'accord sur l'étiquete attribué à n'importe quel échantillon et ***pe*** est l'accord attendu lorsque les deus annotateurs attribuent des étiquettes de manière aléatoire.

L'ensemble des traitements dans la démarche méthodologique est résumé sur le schéma ci-dessous.

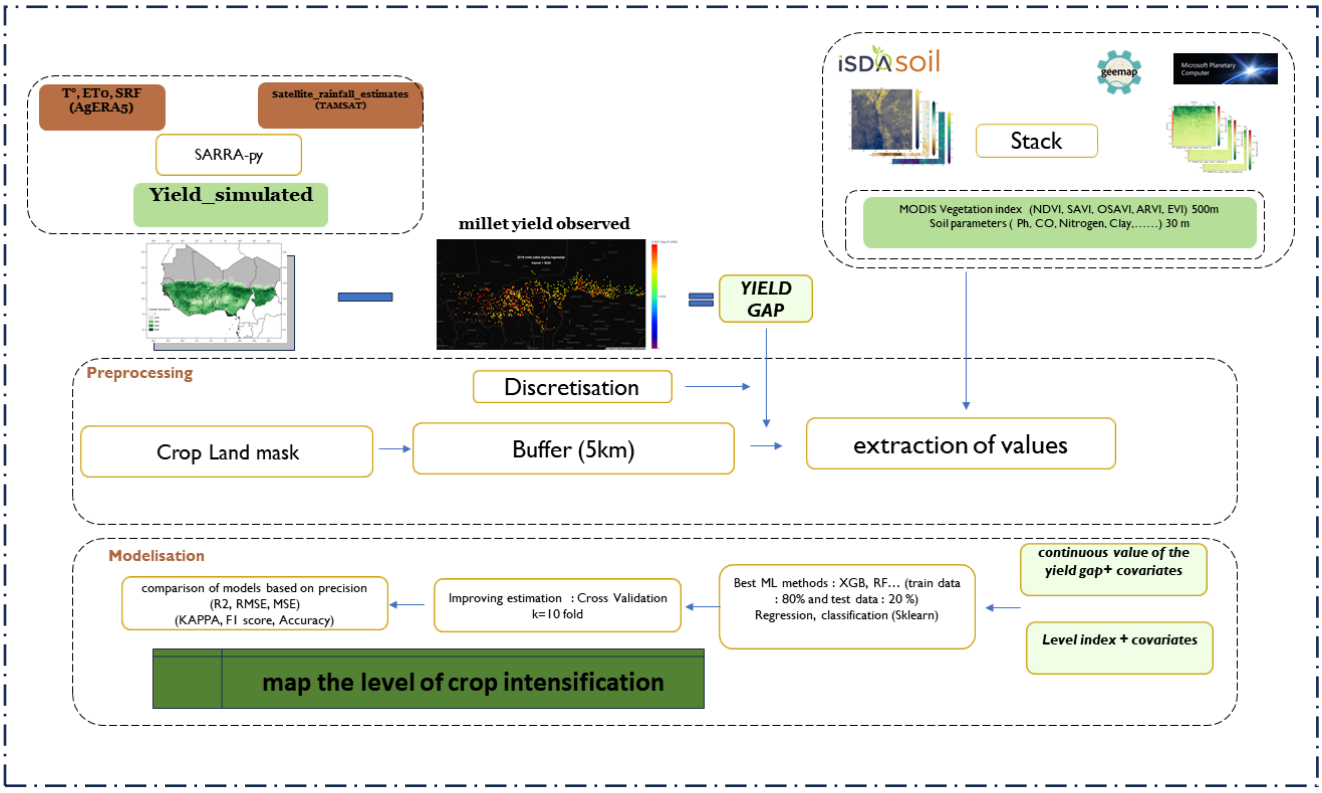


Figure 5: Schéma récapitulatif de la méthodologie

II.8. PRODUCTION DE CARTES DE PREDICTION DE YIELD GAP ET NIVEAU D'INTENSIFICATION

Sur la base des modèles entraînés, nous avons réalisé des inférences à partir des cartes de covariables pour produire des estimations spatialisées des valeurs de yield gap et de niveau d'intensification, sur l'emprise de la zone d'étude à une résolution spatiale de 500m correspondant à la résolution des covaribles issus de MODIS.

CHAPITRE III : RESULTATS

III.1. Simulation de rendement atteignable

Afin de permettre le calcul d'un yield gap, nous avons commencé par réaliser la simulation du rendement atteignable. Dans le processus de modélisation du rendement atteignable du mil dans SARRA-Py, nous avons utilisé les produits d'estimation satellitaire de pluviométrie fournis par TAMSAT (résolution de 0.0375°). Cette résolution définit la résolution de la grille de résultats. Donc, les cartes présentées ci-dessous (Figure 6 et Figure 7) ont une résolution de l'ordre de 4km. Elles représentent les résultats de la simulation des rendements atteignables dans la bande sahélienne pour l'année 2018. Ces rendements varient entre 0 et 6000 kg/ha, avec de fortes valeurs identifiées au centre du Nigéria. Le résultat obtenu avec la méthode "alternative" présente des zones avec des valeurs nulles. Mis à part pour le Burkina Faso et le nord de la bande sahélienne, les rendements retournés par ces deux simulations restent plutôt proches, avec une valeur moyenne de 1930 kg/ha. Dans la suite de cette étude, les rendements atteignables obtenus avec la méthode 'Ru_ISRIC' sont retenus pour le calcul du yield gap. En effet, ce choix se justifie par le fait que nos points d'observation superposés avec les rendement atteignables obtenus avec la méthode alternative se retrouves avec des valeurs nulles.

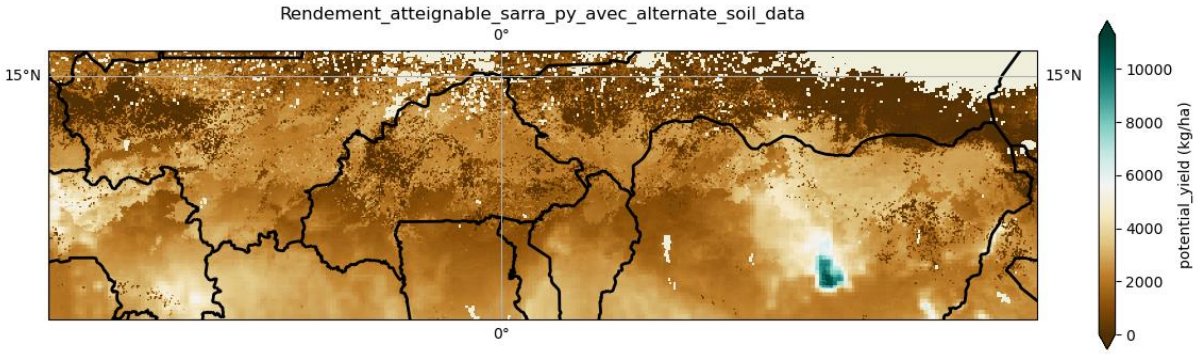


Figure 6: carte de rendement atteignable avec la méthode alternative

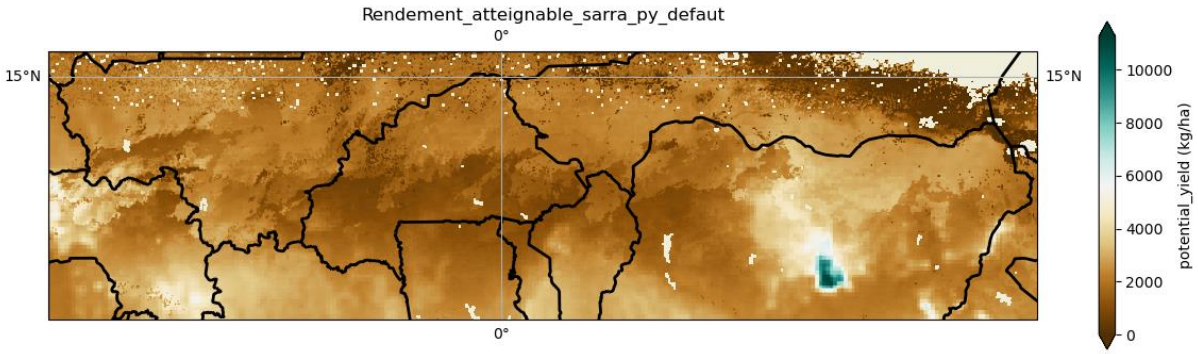


Figure 7: carte de rendement atteignable avec la méthode Ru_ISRIC

III.2. Présentation des valeurs de yield gap

Le yield gap est calculé à partir de points de rendements déclarés dans le jeu de données LSMS-ISA RHoMIS en faisant la différence entre (Ra – Rd). Ce yield gap, qui correspond à un yield gap de niveau II + III (Figure 8), présente des valeurs allant de 3000 kg/ha (cela signifie que la simulation surestime le rendement par rapport au déclaratif, ce qui est attendu car le rendement atteignable simulé est un rendement maximal étant donné les conditions hydriques) à -4000 kg/ha. Les valeurs inférieures ou égales à zéro représentent 5,61% de l'ensemble, ce qui signifie qu'un faible pourcentage des ménages ont un rendement supérieur au rendement atteignable simulé. Ces valeurs sont réparties dans quatre classes de niveau d'intensification d'après la méthode quantile. La figure 9 présente l'histogramme des valeurs pour le yield gap et les boxplots du yield gap au niveau de chaque classe de niveau d'intensification; on observe une distribution bimodale et avec un pic vers 1000 et 2000 kg/ha. La discrétisation en niveaux d'intensification par une approche de quartiles permet de définir 4 classes (élevée, moyenne, faible et très faible) : un niveau d'intensification élevé aura donc une valeur de yield gap faible, et vice versa. Pour la classe 'élevée', le yield gap moyenne est de 306kg/ha ha ; on note cependant un certain nombre de valeurs négatives de yield gap, tombant cependant en outliers de la distribution, ce qui nous conforte dans le choix des critères utilisés pour définir ces classes.

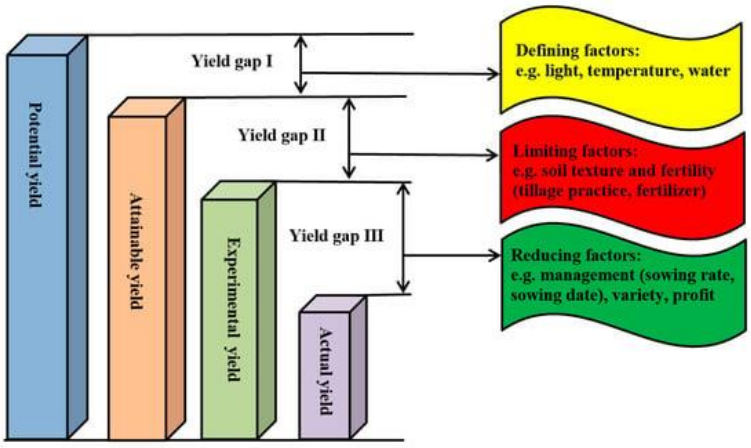


Figure 8: différents niveaux de yield gap, d'après Wang et al 2019

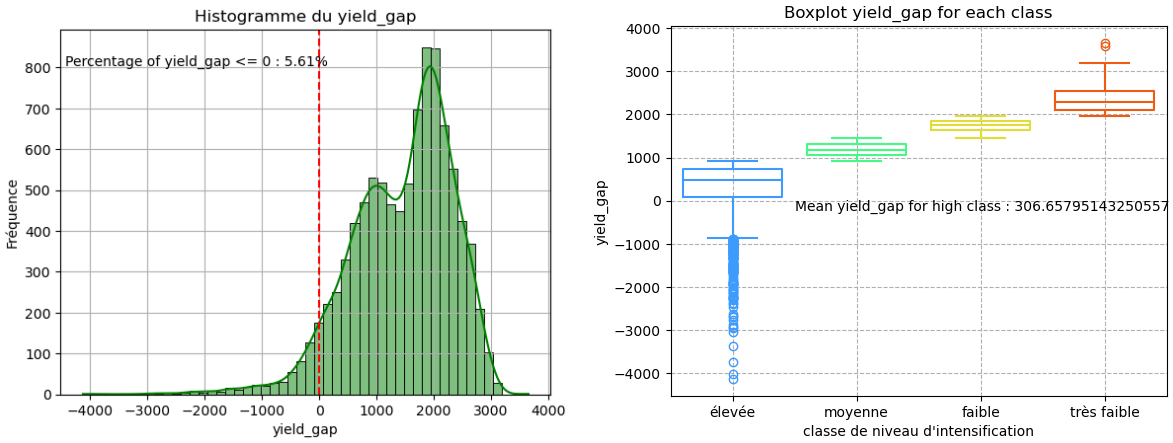


Figure 9: Statistiques du yield gap

III.3. RESULTATS DE LA MODELISATION DU YIELD GAP ET DES NIVEAUX D'INTENSIFICATION PAR MACHINE LEARNING

Dans cette partie, nous présentons les résultats issus de la modélisation par méthodes ML de nos deux variables yield gap et niveau d'intensification. Conformément à la stratégie présentée dans la partie matériel et méthode, nous commençons par présenter les résultats d'un modèle "baseline" avant de présenter les résultats de modèles plus complexes.

III.3.1. Modèle "baseline"

Plusieurs modèles simples dits 'baseline' comportant une seule variable explicative sont implémentés avec l'algorithme random forest (RF), et leur performance sont comparés. Le yield gap étant une variable cible continue, l'exercice à réaliser est une régression, et l'option régression de l'algorithme RF est appliquée. Les résultats obtenus à partir de la validation croisée en k-fold sont récapitulés sur Tableau 2 ci-dessous. En comparant les performances des modèles RF de régression, celles obtenues avec la covariable NDVI sont plus importantes avec un coefficient de détermination R^2 = 71,8%, une RMSE = 529 kg/ha et une MAE = 331,9 kg/ha Figure 10. En procédant à partir du modèle entraîné à l'inférence pour la production d'une carte, on observe des parties de la zone d'étude avec des valeurs de yield gap pouvant atteindre jusqu'à -1000 kg/ha (Figure 11).

Tableau 2: Rapport de modélisation des tests pour la détermination du modèle 'baseline'

				Métriques de performance pour les modèle de régression			Métriques de performance pour les modèles de classification		
Cross_validation	**Model**	**hyperparamètrs**	**Covariables**	**R2**	**RMSE**	**MAE**	**Accuracy (%)**	**F1 score (%)**	**Kappa (%)**
k-fold = 10	Random Forest	N_tree = 200	NDVI	71,80	528,97	331,85	70,27	65	62,35
			LAI	16,29	911,42	714,26	26,32	21,04	25,72
			SRTM	53,01	682,85	479,05	49,41	35	33,57
			CO	42,50	755,39	534,88	52,61	52	34,31
			pH	14,38	921,73	716,09	41,01	43	21,26
			Clay	32,37	819,23	598,18	49,33	53	33,27

Figure 10: Scatterplots des valeurs prédites (prédictions sur les 10 folds à partir du set de test) contre les valeurs observées pour le yield gap, obtenues avec le modèle 'baseline' utilisant NDVI

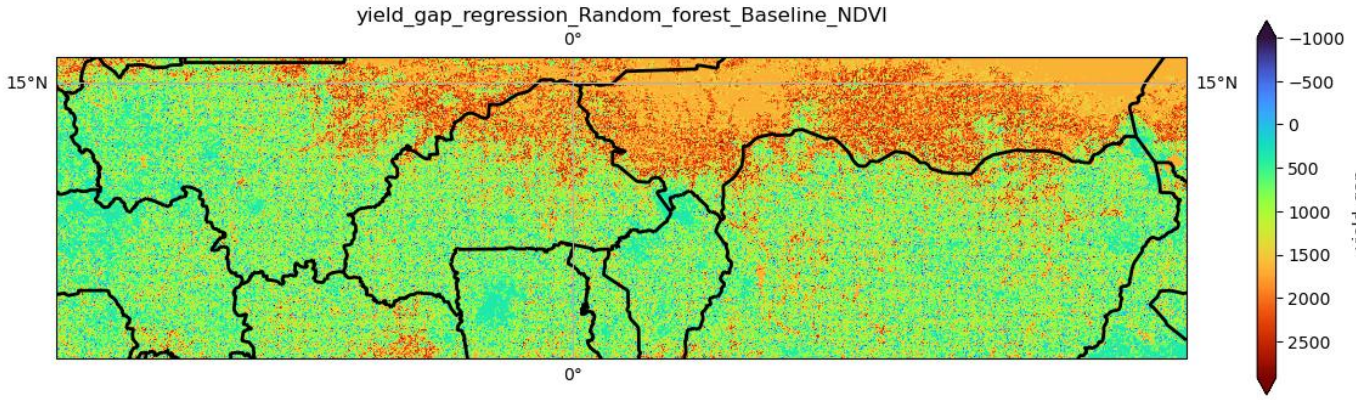


Figure 11: Carte des yield gap du modèle baseline

Nous avons également procédé à l'entraînement de modèles « baseline » de classification en vue de la prédiction de niveaux d'intensification. L'algorithme *RF classifier* est implémenté avec les mêmes covariables que celles utilisés précédemment. Les résultats sont également présentés dans le Tableau 2. Et il s'est encore avéré que le modèle RF entraîné sur la base de la

variable explicative NDVI présente des performances plus élevés avec une accuracy = 70,27%, un F1 score = 65% et un coefficient Kappa = 62,35%. Donc en somme, on peut dire que le NDVI possède une sensibilité par rapport à la modélisation des classes de niveau d'intensification. Ce constat est appuyé par la matrice de confusion (Figure 12) qui montre que le RF avec le NDVI parviennent à modéliser correctement 71% de la classe 'élevée'. La carte produite présente des cohérence, mais nécessite d'être affinée en évaluant les gains de performance potentiels au travers de l'utilisation d'un plus grand nombre de variables explicatives.

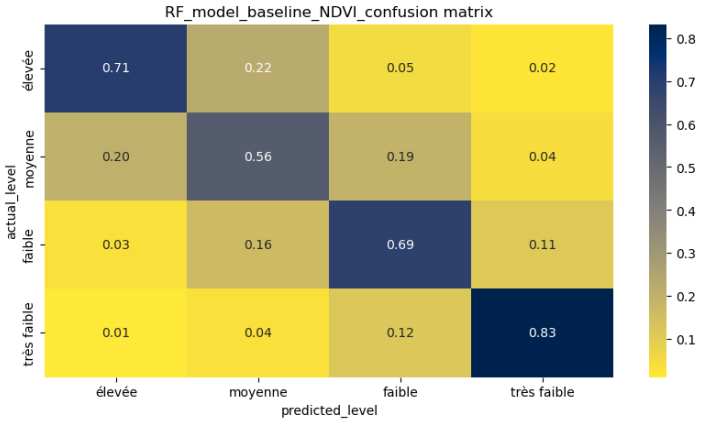


Figure 12: matrice de confusion du modèle RF baseline utilisant la covariable NDVI pour la détermination du niveau d'intensification

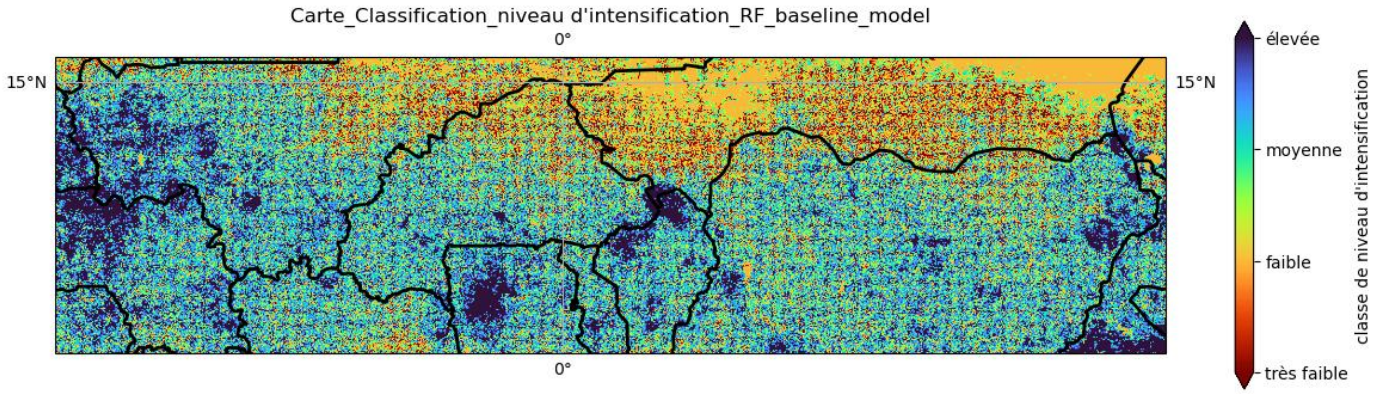


Figure 13 : Carte des niveaux d'intensification obtenues avec le modèle (classification) baseline utilisant NDVI comme variable explicative

III.3.2. Ajout de variables explicatives et comparaison avec d'autres types de modèles de ML

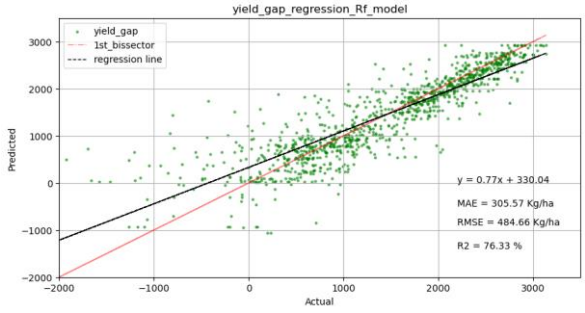
a. Ajout de variables explicatives

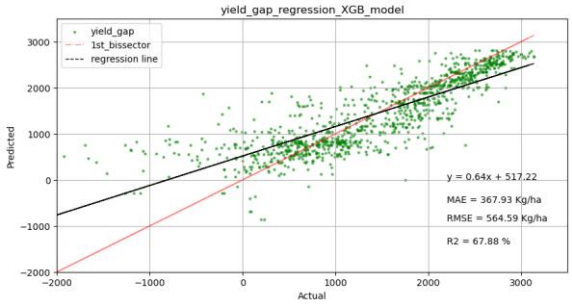
Pour tenter d'augmenter les performances obtenues par le modèle 'baseline', nous partons de l'hypothèse que l'ajout de sources d'informations complémentaires pertinentes par rapport à la

question du yield gap pourrait améliorer les prédictions des modèles. Pour ce faire, d'autres covariables sont stackés avec le NDVI. De plus, d'autres modèles de machine learning sont testés et les hyperparamètres de ces modèles sont conservés (par défaut) en vue de voir l'apport de ces covariables sur la modélisation. Quatre modèles de machine learning (Rf, XGboost, ADboost, Dtree) ont été testés. En plus du RF, le XGB et le DT retournent respectivement des R^2 égaux à 76,33%, 67,88% et 74,21% (Tableau 3). A contrario, le modèle ADB est moins performant avec un R^2 = 42,47%. En ce qui concerne les erreurs générées, le Rf présente les erreurs les moins faibles.

Tableau 3 : Rapport de la régression et de la classification utilisant 6 covariables (Ndvi, Lai, Co, pH, Clay et Srtm)

				Métrique de performance pour les modèle de régression			**Métrique de performance pour les modèle de classification**		
Covariables	**hyperparametres**	**Cross_validation**	**Modèles**	**R2**	**RMSE**	**MAE**	**Accuracy (%)**	**F1 score (%)**	**Kappa (%)**
NDVI, LAI, CO, pH, clay, SRTM	N_tree = 200	k-fold = 10	Random Forest	76,33	484,66	305,57	74,97	77	68,95
			XGboost	67,88	564,59	367,93	72,94	75	65,72
			ADaboost	74,21	505,86	310,44	60,77	62	48,21
	'Defaut'		Decision_tree	42,47	755,58	590,76	70,13	74	31,12





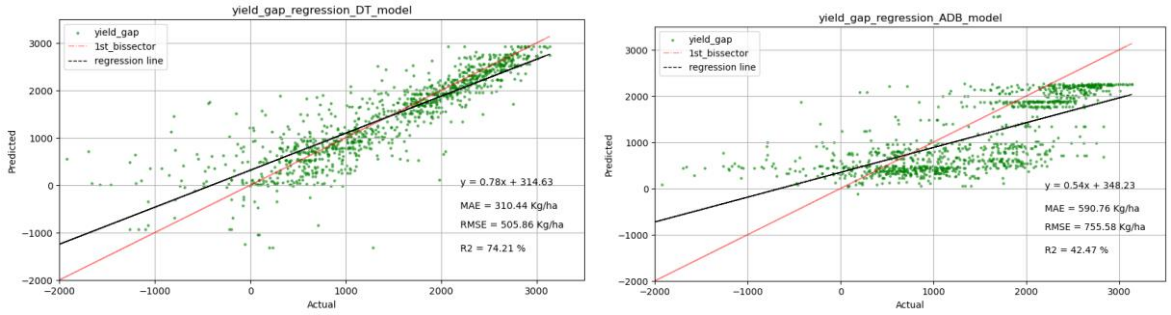


Figure 14: Les scatterplots des valeurs prédites (prédictions sur les 10 folds à partir du set de test) versus les valeurs observées pour le yield gap, obtenues avec les 4 modèles utilisant 6 covariables (Ndvi, Lai, Co, pH, Clay et Srtm)

Au terme du processus d'entraînement des modèles pour la classification des niveaux d'intensification une analyse de *feature importance* est réalisée (figure 15). Les graphiques d'importance des variables montrent d'en plus du NDVI, le carbone organique du sol a une influence importante lors de l'entrainement. Ces graphiques montrent que plus de 50% des yield gap pourrait être dû à la teneur en carbone organique du sol.

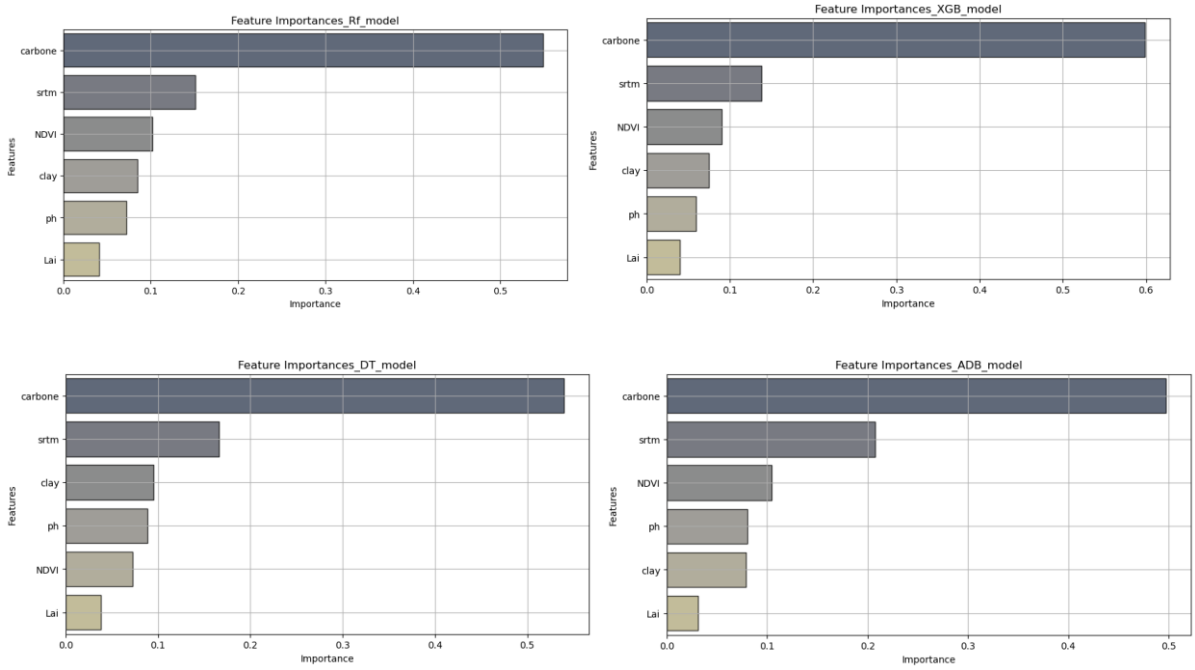


Figure 15: Importance des variables des modèles de régressions du yield gap

Les résultats du modèle RF incluant les 6 covariables montrent des yield_gap de plus de 2000 kg/ha et ce qui traduit un niveau d'intensification extrêmement faible, avec ces valeurs se situant plus particulièrement au sud du Niger, dans la région de Dogo Dogo. Néanmoins, des niveaux d'intensification élevés sont identifiés à l'Ouest du Mali et au nord du Bénin avec des

écarts de rendement pouvant atteindre moins -500 kg/ha. Les cartes issues des autres modèles sont présentées en Annexe (Figure 28, 29 et 30).

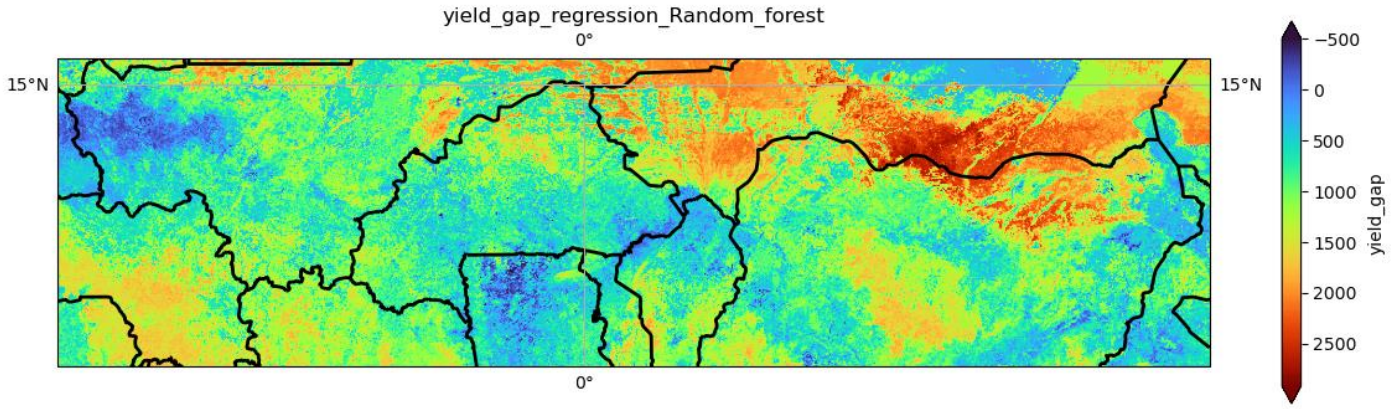
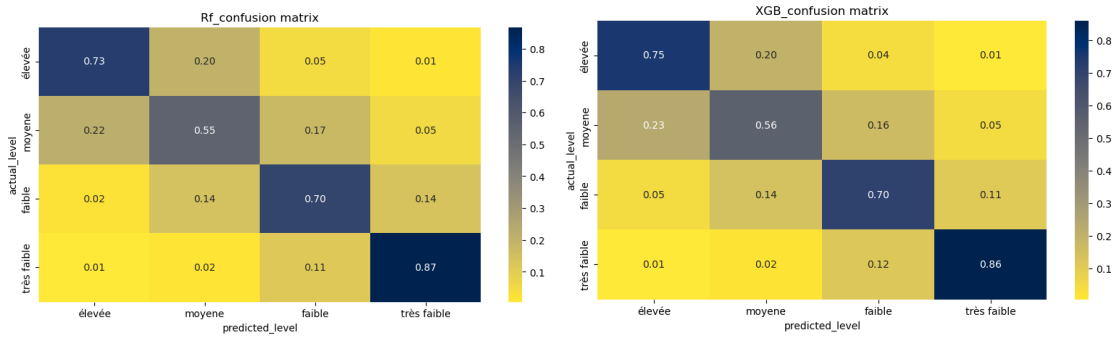


Figure 16: Carte des yield gap obtenue avec le RF_regressor

Une comparaison des performances des quatres modèles est faite et ses résultats sont présentés dans le Tableau3. le modèle RF se montre avec : un coéficient Kappa de 68,95%, un F1 score de 77% et une précision globale de 74,97%. La matrice produite par le RF montre que pour la classe ‘élevée’, 73% des pixels ont été bien classés. Par contre des confusions non négligeables sont notées et peuvent affecter la cartographie des niveaux d’intensification. L’analyse de feature importance (Figure 18) retourne ici des éléments proches et d’autres moins proches par rapport à l’analyse de régression : le NDVI semble jouer sur la classification un role plus important que le carbone. Une fois de plus ici les trois features les plus importantes sont les mêmes que dans l’approche de régression.



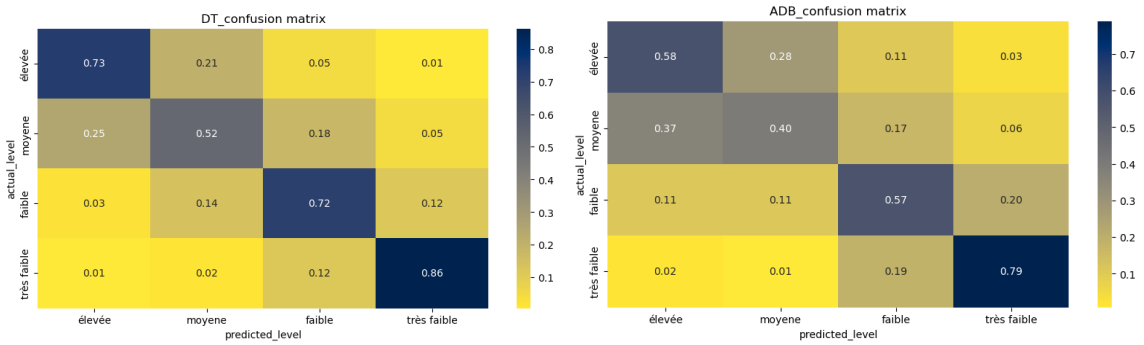


Figure 17: les matrices de confusion des modèles de classification du niveau d'intensification utilisant 6 covariables (Ndvi, Lai, Co, pH, Clay et Srtm)

Les sorties des modèles montrent des niveaux d'intensification fortement influencés par la dynamique de la végétation matérialisé par l'indice NDVI et le carbone organique du sol.

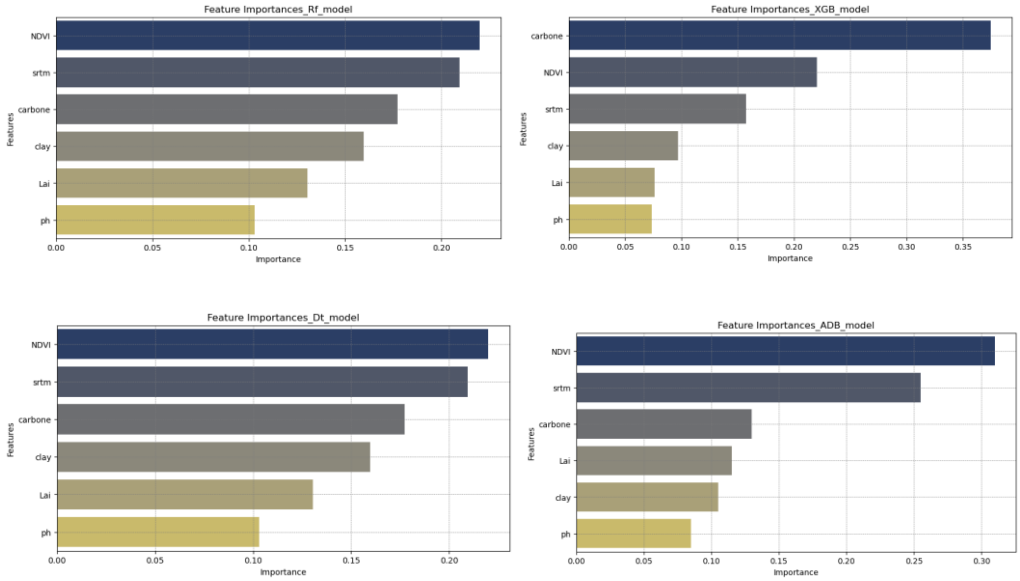


Figure 18: Importance des variables des modèles classification du niveau d'intensification

Le RF étant plus performant, a permis de spatialiser le niveau d'intensification. La carte produite en sorti (figure 19) présente des classes de niveaux d'intensification très faibles à élevées. Cette dernière classe matérialise les zones ou les yield gap sont plus faibles, nuls ou négatifs.

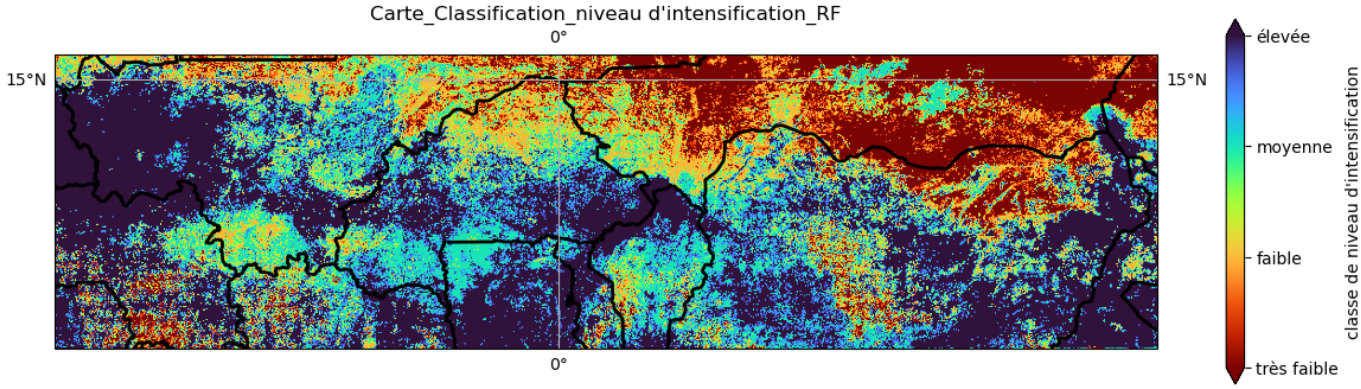
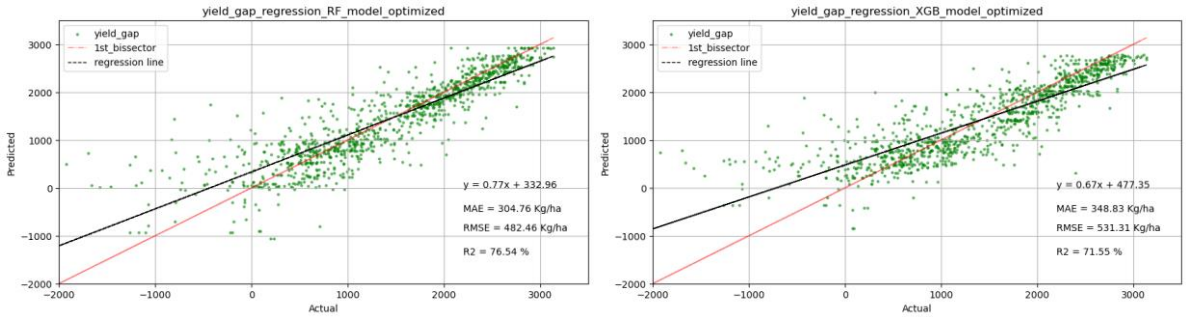


Figure 19: Carte des niveau d'intensification obtenue avec la classification RF utilisant 6 covariables (Ndvi, Lai, Co, pH, Clay et Srtm)

b. Optimisation des modèles

En visualisant la carte de niveau d'intensification ci-dessus, des confusions matricielles sont observées, ce qui peux se justifier par le fait que la variable cible (niveau d'intensification) serait influencé par encore d'autres covariables notamment le teneur en azote du sol pour tenir en compte l'apport de fertilisant azoté pour booster la croissance des cultures, la teneur en argile pour le complexe argilo-humique et la capacité de rétention en eau du sol et le limon (silt) pour tenir en compte la qualité d'aération des sol pour un développement optimal du système racinaire. En outre, des indice de végétation tels que Savi, Msavi, Arvi, Osavi et Evi sont rajouté au Stack pour que les modèles soient sensibles à la dynamique et à la santé de la culture.

Parallèlement, les modèles implémentés ci-dessus sont optimisés en fonctions de leurs hyperparamètres pour gagner plus de précision. Après optimisation, le RF reste le modèle le plus performant avec un R^2 = 76,54% , une RMSE = 482,46Kg/ha, une MAE=304,76Kg/ha.



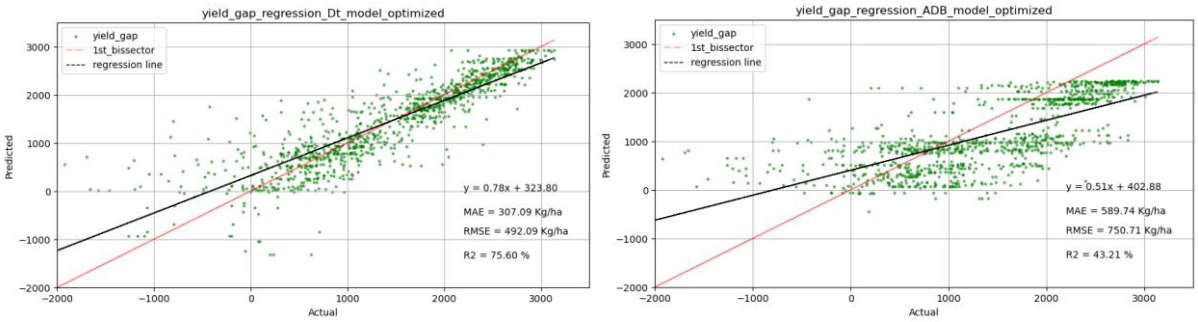


Figure 20: Les scatterplots des valeurs prédites (prédictions sur les 10 folds à partir du set de test) versus les valeurs observées pour le yield gap, obtenues avec les 4 modèles optimisés

En observant les variables plus sensibles au yield gap, le carbone organique du sol ressort avec plus de 30% de valeur d'importance avec le Rf. Il s'en suit l'indice de végétation Arvi et qui avec le XGboost et le ADB influe plus de 20% sur l'estimation du yield gap. Ce qui est plausible car cet indice est très proche du NDVI et est connu pour sa capacité à corriger l'effet atmosphérique.

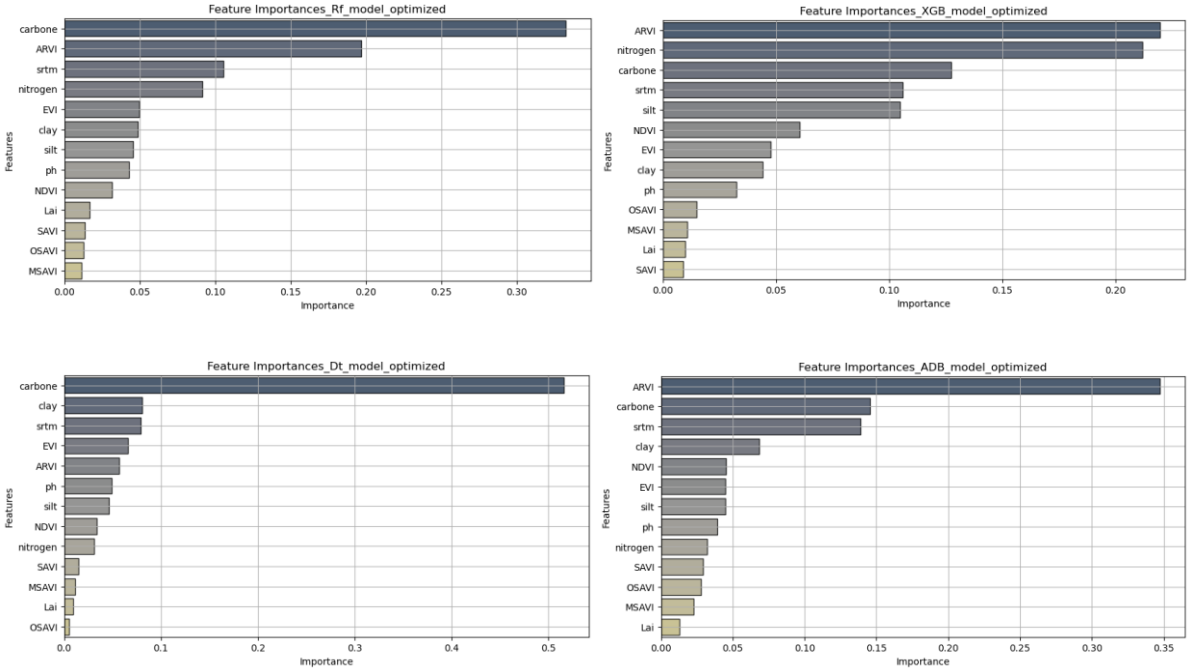


Figure 21: Importance des variables des modèles de régression optimisés

Avec 13 covariables et une optimisation des modèles, la carte du yield gap en sorti présente moins de zone ou la valeur du yield gap est inférieure ou égale à zéro.

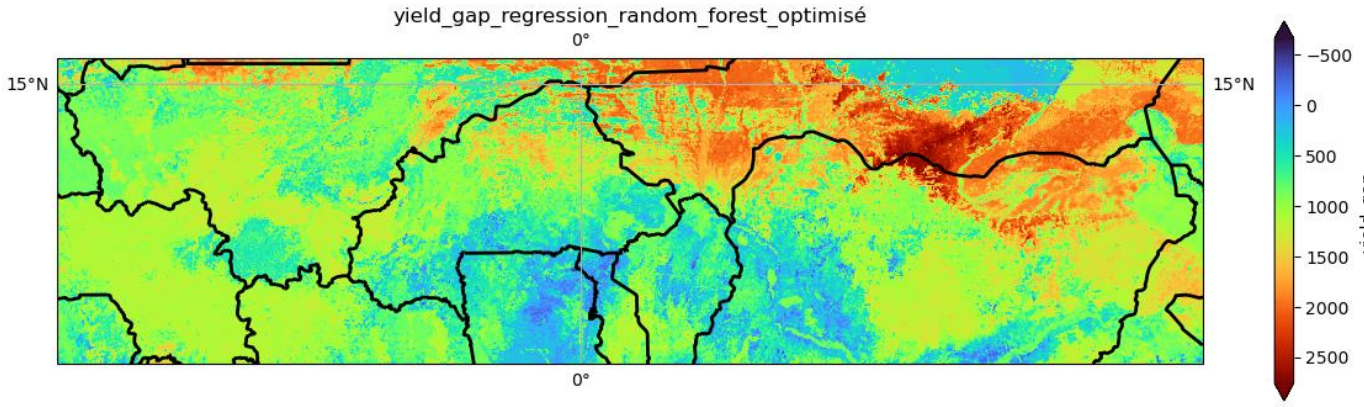


Figure 22: Carte des yield gap avec le Rf optimisé

Les même covariables retenues ci-dessus sont utilisés pour la modélisation du niveau d'intensification. De même, les modèles de classification sont optimisés dans le but d'avoir de meilleures précision. Cette optimisation a permis de passer de 73% à 74% au niveau de la matrice de confusion pour la classe 'élevée' qui est la classe la plus importante représentant le niveau d'intensification le plus élevé. Et toujours dans le dynamique de comparaison, le Rf présente les métriques les plus intéressantes (Tableau 4)

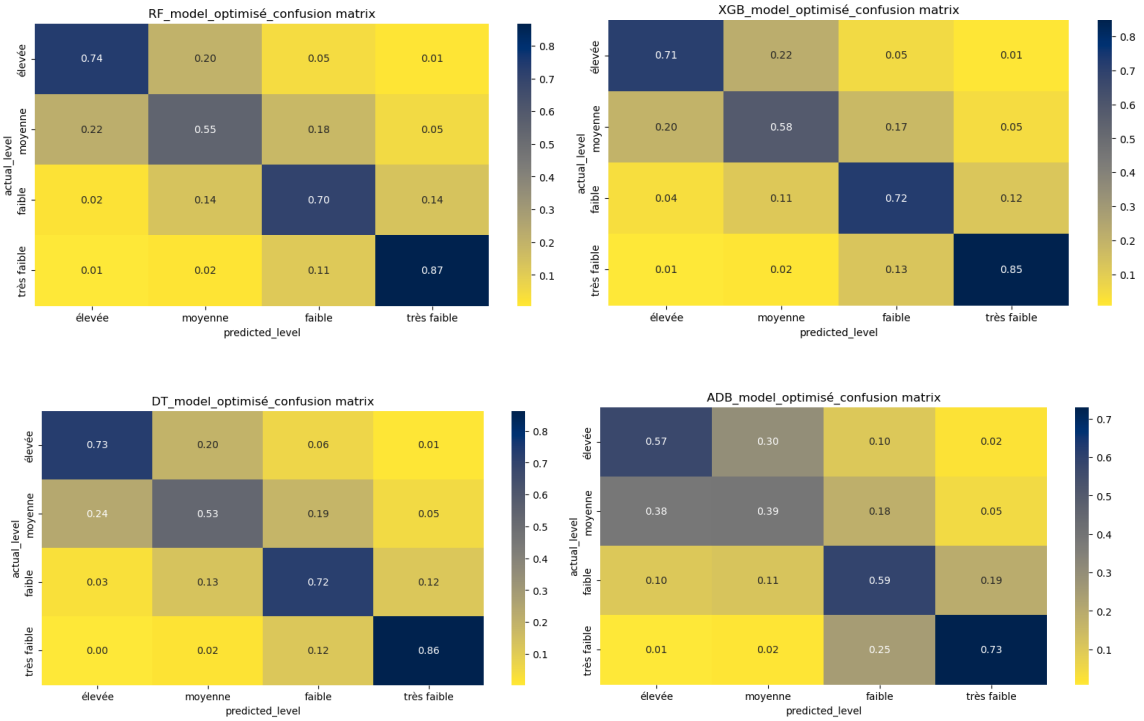


Figure 23 : les matrices de confusion des modèles optimisés de classification du niveau d'intensification

Comme indiqué avec le modèle de régression RF pour estimer le yield gap, la modélisation du niveau d'intensification révèle une sensibilité de l'indice ARVI, suivi du Srtm , du limon (silt), du NDVI et du carbone organique du sol.

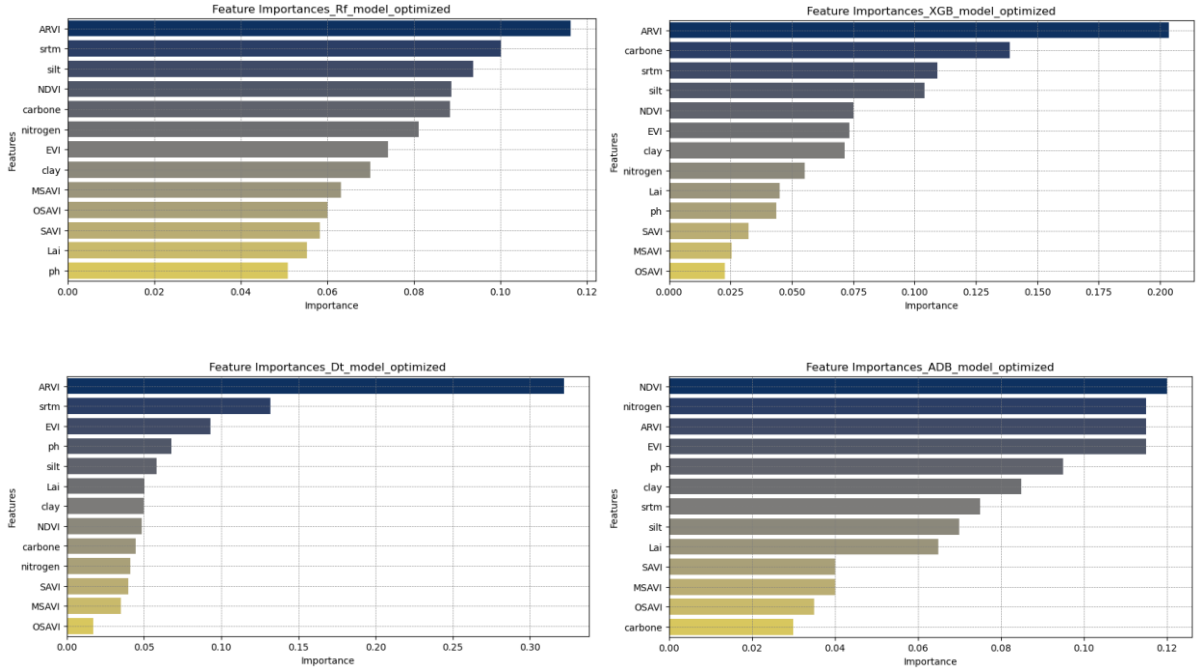


Figure 24: Importance des variables des modèles de classification optimisés

La carte ci-dessous représente la cartographie du niveau d'intensification avec un RF optimisé avec 13 covaribles et des hyperparamètres ajustés. Cette carte est plus complexe car tient en compte plus de facteurs expliquant des niveaux d'intensification plus élevés.

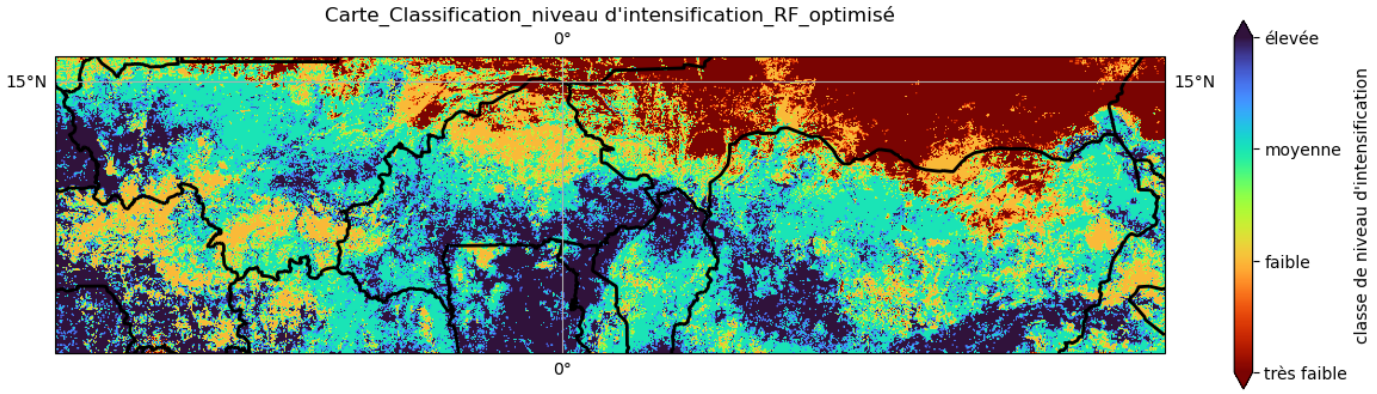


Figure 25: Carte des niveau d'intensification avec le modèle Rf optimisé

Tableau 4 : Rapport de la régression et de la classification

				Métriques de performance pour les modèle de régression			Métriques de performance pour les modèle de classification		
Covariables	**Model_classifieur**	**hyperparametres**	**Cross validation**	**R^2**	**RMSE**	**MAE**	**Accuracy (%)**	**F1 score (%)**	**Kappa (%)**
NDVI, LAI, EVI, OSAVI, MSAVI, N, Silt, SAVI, ARVI, Co, pH, clay, SRTM	Random Forest	N_tree = 300, Md=25, Mf='sqrt', msl=2, mss=10	k-fold = 10	76,54	482,46	304,76	82,32	79	69,13
	XGboost	N_tree = 100, Md=0, Mf='log2', msl=15, mss=100	k-fold = 10	71,55	531,31	348,83	79,97	74	66,72
	ADboost	N_tree = 700, lr=0.001	k-fold = 10	43,21	750,71	589,74	62,45	66	49,31
	Decision_tree	Md = 10, msl=10	k-fold = 10	75,60	492,09	307,09	78,62	74	64,60

Avec :

Md : max_depth, Mf : max_feature, Msl : min_sample_leaf, Mss = min_sample_split, lr : learning_rate

CHAPITRE IV : DISCUSSION

Utilisé principalement en Afrique de l'ouest pour simuler la croissance des cultures, en particulier le mil, le modèle SARRA-H et sa version spatialisée SARRA-Py permet de répondre au besoin spécifique de cette région ou l'agriculture dépend fortement des précipitation qui peuvent être variables d'une année à l'autre. Il détient une capacité éprouvée à simuler le rendement atteignable, c'est-à-dire le rendement que les agriculteurs peuvent raisonnablement espérer atteindre compte tenu des pratiques agricoles locales, des variétés cultivées, et des conditions climatiques spécifiques d'une saison. Dans le cadre de notre étude, ce rendement atteignable varie entre 0 et 6000kg/ha, ce qui s'explique par une forte hétérogénéité des conditions agricoles. (Sultan *et al.*, 2013) ont montré dans leur étude, que ce rendement atteignable est corrélé avec les rendement observés sous une large gamme de cultivars et de pratiques traditionnelles de gestion des cultures.

Cette forte variation des rendements due à l'hétérogénéité des pratiques culturales et des conditions édapho-climatiques sont confirmées par nos résultats du yield gap avec des minimums de -4000kg/ha et des maximums pouvant atteindre jusqu'à 4000kg/ha. Ces importantes valeurs de yield gap sont généralement observé dans des environnements ou les conditions agro climatiques ne sont pas exploitées de manière optimale (Lobell, 2013). Les yield gaps nuls et négatifs, peuvent être interprétés comme des zones où les pratiques agricoles permettent d'atteindre des rendements réels qui surpassent en fait le rendement atteignable tel que modélisé. Ces valeurs très faibles de yield gap pourraient être dû aussi à une sous-estimation des rendements atteignable par SARRA-py. Dans cette étude, nous considérons que ce pourcentage de valeur de yield gap négatif ou nul (5,61%) représente un niveau d'intensification maximale en d'autre terme représente les zones où l'on note une forte potentielle agricole exploitée au maximum. Néanmoins il est à souligner que ce faible pourcentage suggère que la majorité des zones de culture dans la bande sahélienne aurait encore un potentiel non exploité. (Tittonell *et al.*, 2010) stipulent qu'en Afrique subsaharienne les rendements sont souvent bien en deçà de leur potentiel en raison de la gestion suboptimale des ressources. Cette assertion est totalement en phase avec nos valeurs du yield gap.

Cependant, il est intéressant de regarder la répartition spatiale de ce yield gap et de parler de la stratégie déployée pour trouver les facteurs qui influent ce yield gap. Les images MODIS et ses produits dérivés (NDVI, LAI, EVI), nous ont permis de couvrir toute la zones d'étude sur toute la saison végétative de la culture du mil (mai – octobre 2018). Le choix porté sur la moyenne

des indices de végétation est pertinent dans la mesure où il permet de prendre en compte un état physiologique moyen des couverts végétaux sur la saison. (Cheng *et al.*, 2022) affirment que la moyennes des indices de végétation est très corrélée à la croissance des culture ainsi que le rendement réel. En outre, les paramètres physico-chimiques du sol spatialisés issus de ISDA_SOIL tels que le carbone organique, le limon, l'argile, le pH s'avèrent très corréler aux sols fertiles pour l'agriculture et donc en observant les boxplots (Figure 26, annexe), les valeurs des paramètres (CO, silt, clay, N) sont d'autant plus important que les niveau d'intensification sont élevée. Ainsi, Les indices de végétation et les paramètres physico-chimiques cités ont servi de covariables pour la spatialisation du yield gap et pour la modélisation du niveau d'intensification avec des modèle de machine learning. Le résultat obtenu avec le modèle baseline calibré avec le NDVI et le RF s'est montré plus performant que les autres modèles RF obtenus avec d'autres covariables, avec un R^2=71% pour spatialiser le yield gap et une accuracy=70% pour modéliser le niveau d'intensification. Ce qui peut s'expliquer par le fait que le NDVI, étant un indicateur clé de la vigueur et de la densité des cultures, est très corrélé à la biomasse photosynthétique des cultures et ces valeurs sont les plus précises pendant la saison agricole, notamment lors du stade de croissance active de la culture (Mousavi *et al.*, 2024). De plus, comme le montre la Figure 27 (annexe) la moyenne de NDVI est corrélée linéairement de 55% avec le niveau d'intensification, donc le NDVI reste une covariable intéressante pour sa modélisation.

Cependant, en ajoutant d'autres covariables (Co, LAI, Srtm, Ph, et Clay) pour la spatialisation et la modélisation avec 4 modèles de machine learning, les résultats montrent que le RF est plus performant que les autres modèle avec R^2=76,33%, une RMSE=484,66 Kg/ha et une MAE=305,57 pour spatialiser le yield gap, et un coefficient Kappa de 68,95%, un F1 score de 77% et une précision globale de 74,97% pour la modélisation du niveau d'intensification. En effet étant plus performant devant le XBoost, le *DT et* ADaboost pour spatialiser le yield gap et pour modéliser le niveau d'intensification, le RF présente une robustesse faces aux données hétérogènes et bruitées, ce qui est le cas dans notre étude avec l'utilisation combinée des indices de végétation et des paramètres physico-chimiques du sol. En outre le RF est reconnu pour sa performance élevée en termes de précision et de généralisation. Grâce à la méthode d'agrégation des résultats de nombreux arbres de décision, il est moins susceptible de surapprendre sur des jeux de données complexes, ce qui est essentiel pour une bonne prédiction spatiale à grande échelle (Jeong *et al.*, 2016).

Ainsi, le carbone organique du sol, le NDVI et le SRTM sont les covariables avec les valeurs d'importances les plus élevées pour tous les modèles aussi bien pour la régression du yield gap que sur la classification du niveau d'intensification. (Pettorelli *et al.,* 2014) qui stipule que le NDVI est couramment utilisés pour évaluer la santé et la vigueur de la végétation, et leur variabilité peut être révélatrice de différences dans les stades de croissance des cultures, la santé ou même la mise en œuvre de diverses pratiques agricoles. En outre la forte valeur d'importance du carbone organique lors de la régression ou de la classification confirme notre hypothèse lors des choix des covariables: le Co est une clé de fertilité des sol et son abondance dans le sol peut être favoriser par des amendements organiques. Ces résultats sont en phase avec les travaux de (Mousavi *et al.,* 2024) qui stipule que l'application du Co et d'engrais organique peut entraîner une augmentation de la capacité de rétention d'eau, de la porosité du sol, de la stabilité des agrégats et une diminution du compactage du sol et de la formation de croûtes de surface, ce qui peut se traduit par un niveau d'intensification élevé et par conséquent entraîner une production agricole élevée. De plus le Co est un substrat pour les micro-organismes du sol, et la disponibilité de ce Co influence la diversité et l'activité des microbes du sol (Fan *et al.*, 2005), (Kanchikerimath and Singh, 2001).

En plus de ces résultats, nous en avons produit encore d'avantage avec plus de covariables (Osavi, Msavi, Evi, Savi , Arvi, Silt et Nitrogen) et avec des hyperparamètres ajustés dans le but d'obtenir des modèles plus optimisés et plus complexifiés. Cette stratégie nous a permis de gagner en précision avec toujours le RF plus performant avec un R^2=76,54% pour le yield gap et 82,32% d'accuracy pour le niveau d'intensification. En outre ces modèles optimisés retournent pour le RF du yield gap une valeur d'importance pour le Co plus élevée suivi de l'indice ARVI, le SRTM et de l'azote. Tandis que pour le niveau d'intensification les covariables qui ressortent sont Arvi, Srtm, Silt et Ndvi. Il faut noter que l'indice Arvi réagit de manière semblable au NDVI sur des données de réflectances spectrales collectées à la surface de la terre comme le rendement agricole (Martin Ledant, 2007). Il demeure un indice de végétation relativement insensible aux facteurs atmosphériques tels que les aérosols et corrige l'indice de végétation NDVI en atténuant les effets de diffusion atmosphérique. Par conséquent on peut dire qu'il est aussi un indicateur de la santé des cultures tout en corrigeant les effets atmosphériques. en plus du NDVI, du ARVI, et du carbone organique du sol, le limon (Silt) et l'azote (N) influent fortement sur le yielg gap et le niveau d'intensification. En ce qui concerne l'azote total du sol, il est l'élément responsable de la croissance des cultures et donc son déficit entraine des malades abiotiques comme la chlorose (carence azotée) et par conséquent des

rendements très faible car le stade de croissance de la plante sera retardé. (Grzebisz and Łukowiak, 2021) révèlent dans leur étude que l'apport insuffisant d'azote à une plante pendant ses étapes cardinales de formation du rendement résulte de deux variabilités majeures, la première est la variabilité spatiale des caractéristiques du sol responsables de l'approvisionnement en eau d'une plante, servant également de transporteur de nutriments. La seconde est une variabilité verticale des facteurs du sol, déterminants pour les réserves de nutriments disponibles et leur accessibilité en saison pour la culture. De plus ils affirment que le diagnostic de l'état de fertilité du sol est important pour le développement de techniques d'application de l'azote et des nutriments, soutenant son efficacité d'utilisation pour assurer une production optimale. Ce diagnostic requiert aussi une bonne connaissance des sols et des techniques comme la modulation intra parcellaire pour mieux optimiser les apports afin de conserver ou de restaurer le niveau de fertilité des sols. Cet apport de dose optimale au bon moment et au bon endroit en amendement organique ou minérale ainsi que pour l'irrigation avec l'utilisation de la technologie (agriculture de précision) se traduit par un niveau d'intensification plus élevé et impactera directement le rendement agricole au niveau des pays situés dans la bande sahélienne.

Cependant, les cartes de yield gap et de niveau d'intensification montrent des zones assez alarmants comme au sud du Niger vers les village de Dogo Dogo avec des yield gap de plus de 1000kg/ha et des niveau d'intensification très faible. En effet, c'est des zones ou les producteurs apportent de l'eau au seau ou à la calebasse à partir de puisards (Patrick Delmas, 2012). Ce déficit technologique explique les rendements très faibles et c'est aussi le cas dans quasiment beaucoup de zones surtout au niveau des petits producteurs ou dans les exploitation familiales. Néanmoins, les cartes de yield gap et de niveau d'intensification montrent des zones avec un grand potentiel agricole comme à proximité du fleuve le Niger et du barrage de manantali au Mali. Ce qui est complétement cohérent car il y aura une absence totale de déficit hydrique. (B.Sultan et al, 2020) stipulent qu'avec des apports de fertilisants, comme la micro-fertilisation, et la réduction du déficit hydrique à la parcelle, les rendements des cultures de décrue peuvent être très nettement améliorés. Le rendement du sorgho de décrue arrosé et fertilisé peut dépasser plusieurs tonnes par hectare pour les parcelles à proximité du barrage. En outre, nos cartes montre un niveau d'intensification très élevé au nord du Bénin ou il on note des pratiques de moto-mécanisation agricole. Aujourd'hui, la quasi-totalité des superficies au nord du pays cultivées est labourée avec les tracteurs et la quasi-totalité des exploitations agricoles a recours au labour motorisé (Taupin, 2023). Donc nous pouvons dire que la méthodologie déployée dans

cette étude pour modéliser des niveau d'intensification est assez robuste car ayant même la possibilité de capter des pratiques agricoles qui traduit un niveau d'intensification intense telle que la moto-mécanisation agricole.

CONCLUSION ET PERSPECTIVES

En somme, cette étude vient éclaircir le débat sur les facteurs potentiels qui influencent les écart de rendement et qui expliquent les niveau d'intensification en Afrique subsaharienne. Cette analyse s'est basé principalement sur les paramètres physico-chimique du sol, les indices de végétation issus de la télédétection et les approches de machine learning. Les résultats ont montré que les attributs du sol peuvent affecter considérablement la variabilité du yield gap, et que la nécessité de surveiller et de suivre ces attributs à l'aide de données de télédétection est essentielle pour étudier le yield gap. Cependant, les indices de végétation et surtout la disponibilité des données de sol en format raster dans l'API ISDA SOIL a permis d'aller plus loin dans l'étude des yield gap et de la modélisation du niveau d'intensification agricole en Afrique subsaharienne. (Khechba *et al.*, 2021) révèlent dans sa revue que les articles qui ont étudié le yield gap avec l'utilisation des attributs du sol sont rares ou inexistants. De plus ils affirme que l'absence de cartes numériques des sols dérivées de la télédétection à haute résolution (c'est-à-dire des cartes de fertilité) peut expliquer en partie la mauvaise utilisation des propriétés du sol par les chercheurs africains pour les études d'analyse des écarts de rendement. Donc oui, l'API ISDA SOIL nous a permis d'étudier l'influence des éléments comme le carbone organique du sol sur le yield gap et/ou le niveau d'intensification agricole. Aussi, l'approche machine learning dans cette étude s'est montrée aussi très efficaces malgré la qualité de nos données. Donc les cartes de yield gap et de niveau d'intensification produites peuvent permettre aux décideurs de revoir les politiques à mener dans certaines zones où le niveau d'intensification est très faible en vue d'éradiquer le phénomène d'insécurité alimentaire qui sévit en Afrique de l'ouest. Pour se faire, il va falloir réduire le déficit technologique et moderniser les pratiques agricoles en adoptant celles de l'agriculture de précision. La modulation intra parcellaire et l'irrigation au point avec l'utilisation des capteurs (sondes capacitives) permettront d'améliorer la fertilité des sol, les restaurés et optimiser les intrants agricoles.

Cependant il est à noter que seules les facteurs agronomiques ne sont pas responsables des yield gap élevé ou les niveau d'intensification faible. Les facteurs socio-économiques influencent fortement aussi le niveau d'intensification agricole. Donc il sera pertinent à l'avenir de tenir en compte ces facteurs socio-économiques tels que la densité de la population, la composante genre, et le pourcentage des jeunes dans les champs agricoles. De plus, l'accès au intrants est aussi un véritable problème surtout pour les petits producteurs et par conséquent se retrouvent dans des campagnes agricoles sans amendement ni fertilisant. L'accès aux assurances et aux

financements agricoles sont fortement lié à la problématique du foncier et des conflits entre les agro-business et les petits producteurs (agriculture familiales). Tous ces facteurs pourraient à l'avenir être tenu en compte pour une pareille étude à moins qu'ils soient disponibles en données de télédétection ce qui est n'est pas évident.

D'autres covariables tels que les métriques phénologiques (SOS, EOS, POS…) peuvent aussi être étudier pour tenir en compte les dates de semis, la variété des cultures…

BIBLIOGRAPHIE

Affholder, F. (1997) 'Empirically modelling the interaction between intensification and climatic risk in semiarid regions', Field Crops Research, 52(1), pp. 79–93. Available at: https://doi.org/10.1016/S0378-4290(96)03453-3.

Amgain, L.P. et al. (2021) 'Nutrient expert® rice - an alternative fertilizer recommendation strategy to improve productivity, profitability and nutrient use efficiency of rice in Nepal', Journal of Plant Nutrition, 44(15), pp. 2258–2273. Available at: https://doi.org/10.1080/01904167.2021.1889590.

Barboza, T.O.C. et al. (2023) 'Performance of Vegetation Indices to Estimate Green Biomass Accumulation in Common Bean', AgriEngineering, 5(2), pp. 840–854. Available at: https://doi.org/10.3390/agriengineering5020052.

Bégué, A. et al. (2023) 'How Well Do EO-Based Food Security Warning Systems for Food Security Agree? Comparison of NDVI-Based Vegetation Anomaly Maps in West Africa', IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 16, pp. 1641–1653. Available at: https://doi.org/10.1109/JSTARS.2023.3236259.

Cassman, K.G. (2012) 'What do we need to know about global food security?', Global Food Security, 1(2), pp. 81–82. Available at: https://doi.org/10.1016/j.gfs.2012.12.001.

Cheng, E. et al. (2022) 'Wheat yield estimation using remote sensing data based on machine learning approaches', Frontiers in Plant Science, 13. Available at: https://doi.org/10.3389/fpls.2022.1090970.

Dehkordi, P.A. et al. (2020) 'Yield Gap Analysis Using Remote Sensing and Modelling Approaches: Wheat in the Northwest of Iran', International Journal of Plant Production, 14(3), pp. 443–452. Available at: https://doi.org/10.1007/s42106-020-00095-4.

Earth Science Data Systems, N. (2024) MODIS | Earthdata. Earth Science Data Systems, NASA. Available at: https://www.earthdata.nasa.gov/sensors/modis (Accessed: 20 June 2024).

Fan, T. et al. (2005) 'Long-term fertilization effects on grain yield, water-use efficiency and soil fertility in the dryland of Loess Plateau in China', Agriculture, Ecosystems & Environment, 106(4), pp. 313–329. Available at: https://doi.org/10.1016/j.agee.2004.09.003.

Gerber, J.S. et al. (2024) 'Global spatially explicit yield gap time trends reveal regions at risk of future crop yield stagnation', Nature Food, 5(2), pp. 125–135. Available at: https://doi.org/10.1038/s43016-023-00913-8.

Grzebisz, W. and Łukowiak, R. (2021) 'Nitrogen Gap Amelioration Is a Core for Sustainable Intensification of Agriculture—A Concept', Agronomy, 11(3), p. 419. Available at: https://doi.org/10.3390/agronomy11030419.

Gumma, M.K. et al. (2024) 'Optimizing Crop Yield Estimation through Geospatial Technology: A Comparative Analysis of a Semi-Physical Model, Crop Simulation, and Machine Learning Algorithms', AgriEngineering, 6(1), pp. 786–802. Available at: https://doi.org/10.3390/agriengineering6010045.

Hisse, I.R. et al. (2022) 'Annual productivity of cropping sequences: Responses to increased intensification levels', European Journal of Agronomy, 137, p. 126506. Available at: https://doi.org/10.1016/j.eja.2022.126506.

Hou, J. et al. (2019) 'Characteristics of vegetation activity and its responses to climate change in desert/grassland biome transition zones in the last 30 years based on GIMMS3g', Theoretical and Applied Climatology, 136(3), pp. 915–928. Available at: https://doi.org/10.1007/s00704-018-2527-0.

Jeong, J.H. et al. (2016) 'Random Forests for Global and Regional Crop Yield Predictions', PLOS ONE, 11(6), p. e0156571. Available at: https://doi.org/10.1371/journal.pone.0156571.

Kanchikerimath, M. and Singh, D. (2001) 'Soil organic matter and biological properties after 26 years of maize–wheat–cowpea cropping as affected by manure and fertilization in a Cambisol in semiarid region of India', Agriculture, Ecosystems & Environment, 86(2), pp. 155–162. Available at: https://doi.org/10.1016/S0167-8809(00)00280-2.

Khechba, K. et al. (2021) 'Monitoring and Analyzing Yield Gap in Africa through Soil Attribute Best Management Using Remote Sensing Approaches: A Review', Remote Sensing, 13(22), p. 4602. Available at: https://doi.org/10.3390/rs13224602.

Kizilgeci, F. et al. (2021) 'Normalized Difference Vegetation Index and Chlorophyll Content for Precision Nitrogen Management in Durum Wheat Cultivars under Semi-Arid Conditions', Sustainability, 13(7), p. 3725. Available at: https://doi.org/10.3390/su13073725.

Li, C. et al. (2022) 'Maize Yield Estimation in Intercropped Smallholder Fields Using Satellite Data in Southern Malawi', Remote Sensing, 14(10), p. 2458. Available at: https://doi.org/10.3390/rs14102458.

Lobell, D.B. (2013) 'The use of satellite data for crop yield gap analysis', Field Crops Research, 143, pp. 56–64. Available at: https://doi.org/10.1016/j.fcr.2012.08.008.

Lobell, D.B., Cassman, K.G. and Field, C.B. (2009) 'Crop Yield Gaps: Their Importance, Magnitudes, and Causes', Annual Review of Environment and Resources, 34(Volume 34, 2009), pp. 179–204. Available at: https://doi.org/10.1146/annurev.environ.041008.093740.

LSMS-ISA (no date) World Bank. Available at: https://www.worldbank.org/en/programs/lsms/initiatives/lsms-ISA (Accessed: 15 July 2024).

Mousavi, S.R. et al. (2024) 'Spatial prediction of winter wheat yield gap: agro-climatic model and machine learning approaches', Frontiers in Plant Science, 14. Available at: https://doi.org/10.3389/fpls.2023.1309171.

Paliwal, A. et al. (2022) 'Using Microsatellite Data to Map the Persistence of Field-level Yield Gaps and Their Drivers in Smallholder Systems'. Available at: https://doi.org/10.21203/rs.3.rs-1654249/v1.

Parra, G., Borrás, L. and Gambin, B.L. (2022) 'Crop attributes explaining current grain yield dominance of maize over sorghum', Field Crops Research, 275, p. 108346. Available at: https://doi.org/10.1016/j.fcr.2021.108346.

PARTIE 1 Les systèmes d'alerte précoce (SAP) (2008) msf-crash.org. Available at: https://msf-crash.org/fr/publications/laide-alimentaire-et-la-politique-des-chiffres-en-ethiopie-2002-2004/partie-1-les (Accessed: 9 July 2024).

Pettorelli, N. et al. (2014) 'Satellite remote sensing for applied ecologists: opportunities and challenges', Journal of Applied Ecology, 51(4), pp. 839–848. Available at: https://doi.org/10.1111/1365-2664.12261.

Remote Sensing | Free Full-Text | Maize Yield Estimation in Intercropped Smallholder Fields Using Satellite Data in Southern Malawi (no date). Available at: https://www.mdpi.com/2072-4292/14/10/2458 (Accessed: 24 June 2024).

Shuai, G. and Basso, B. (2022) 'Subfield maize yield prediction improves when in-season crop water deficit is included in remote sensing imagery-based models', Remote Sensing of Environment, 272, p. 112938. Available at: https://doi.org/10.1016/j.rse.2022.112938.

Sultan, B. et al. (2013) 'Assessing climate change impacts on sorghum and millet yields in the Sudanian and Sahelian savannas of West Africa', Environmental Research Letters, 8(1), p. 014040. Available at: https://doi.org/10.1088/1748-9326/8/1/014040.

Sultan, B., Defrance, D. and Iizumi, T. (2019) 'Evidence of crop production losses in West Africa due to historical global warming in two crop models', Scientific Reports, 9(1), p. 12834. Available at: https://doi.org/10.1038/s41598-019-49167-0.

Tadele, Z. (2017) 'Raising Crop Productivity in Africa through Intensification', Agronomy, 7, p. 22. Available at: https://doi.org/10.3390/agronomy7010022.

Takoutsing: An assessment of the variation of soil... - Google Scholar (no date). Available at: https://scholar.google.com/scholar_lookup?title=An+assessment+of+the+variation+of+soil+properties+with+landscape+attributes+in+the+highlands+of+Cameroon&author=Takoutsing,+B.&author=Weber,+J.C.&author=Martin,+J.A.R.&author=Shepherd,+K.&author=Aynekulu,+E.&author=Sila,+A.&publication_year=2018&journal=Land+Degrad.+Dev.&volume=29&pages=2496%E2%80%932505&doi=10.1002/ldr.3075 (Accessed: 19 June 2024).

Taupin, M. (2023) Quels enjeux pour l'agriculture du nord Benin face aux dynamiques de moto-mécanisation ? : diagnostic agraire de l'arrondissement d'Ina dans le département du Borgou. other. CIRAD - UMR Innovation, 389 avenue Agropolis, 34980 Montferrier-sur-Lez. Available at: https://dumas.ccsd.cnrs.fr/dumas-04428595 (Accessed: 13 August 2024).

Teboh, J.M. et al. (2011) 'Applicability of Ground-based Remote Sensors for Crop N Management in Sub Saharan Africa', Journal of Agricultural Science, 4(3), p. p175. Available at: https://doi.org/10.5539/jas.v4n3p175.

TFE - Les indices de végétation (2015) SlideShare. Available at: https://fr.slideshare.net/MartinLedant/tfe-les-indices-de-vgtation (Accessed: 12 August 2024).

Tittonell, P. et al. (2010) 'The diversity of rural livelihoods and their influence on soil fertility in agricultural systems of East Africa – A typology of smallholder farms', Agricultural Systems, 103(2), pp. 83–97. Available at: https://doi.org/10.1016/j.agsy.2009.10.001.

Traore, A. (2022) Changement climatique et agriculture en Afrique subsaharienne. Perception des agriculteurs et impact de l'association entre une céréale et une légumineuse sur les rendements des deux espèces et leur variabilité inter-annuelle sous climat actuel et futur. Cas du sorgho et du niébé dans l'environnement soudano-sahélien. phdthesis. Sorbonne Université. Available at: https://theses.hal.science/tel-03847646 (Accessed: 15 July 2024).

Vintrou, E. (no date) 'Cartographie et caractérisation des systèmes agricoles au Mali par télédétection à moyenne résolution spatiale'.

Wang, J. et al. (2023) 'Consistency and uncertainty of remote sensing-based approaches for regional yield gap estimation: A comprehensive assessment of process-based and data-driven models', Field Crops Research, 302, p. 109088. Available at: https://doi.org/10.1016/j.fcr.2023.109088.

Wu, S. et al. (2020) 'Spatial and Temporal Changes in the Normalized Difference Vegetation Index and Their Driving Factors in the Desert/Grassland Biome Transition Zone of the

Sahel Region of Africa', Remote Sensing, 12(24), p. 4119. Available at: https://doi.org/10.3390/rs12244119.

Zhao, Y. et al. (2015) 'Using satellite remote sensing to understand maize yield gaps in the North China Plain', Field Crops Research, 183, pp. 31–42. Available at: https://doi.org/10.1016/j.fcr.2015.07.004.

Zida, W.A., Bationo, B.A. and Waaub, J.-P. (2020) 'Re-greening of agrosystems in the Burkina Faso Sahel: greater drought resilience but falling woody plant diversity', Environmental Conservation, 47(3), pp. 174–181. Available at: https://doi.org/10.1017/S037689292000017X.

Zsebő, S. et al. (2024) 'Yield Prediction Using NDVI Values from GreenSeeker and MicaSense Cameras at Different Stages of Winter Wheat Phenology', Drones, 8(3), p. 88. Available at: https://doi.org/10.3390/drones8030088.

ANNEXE

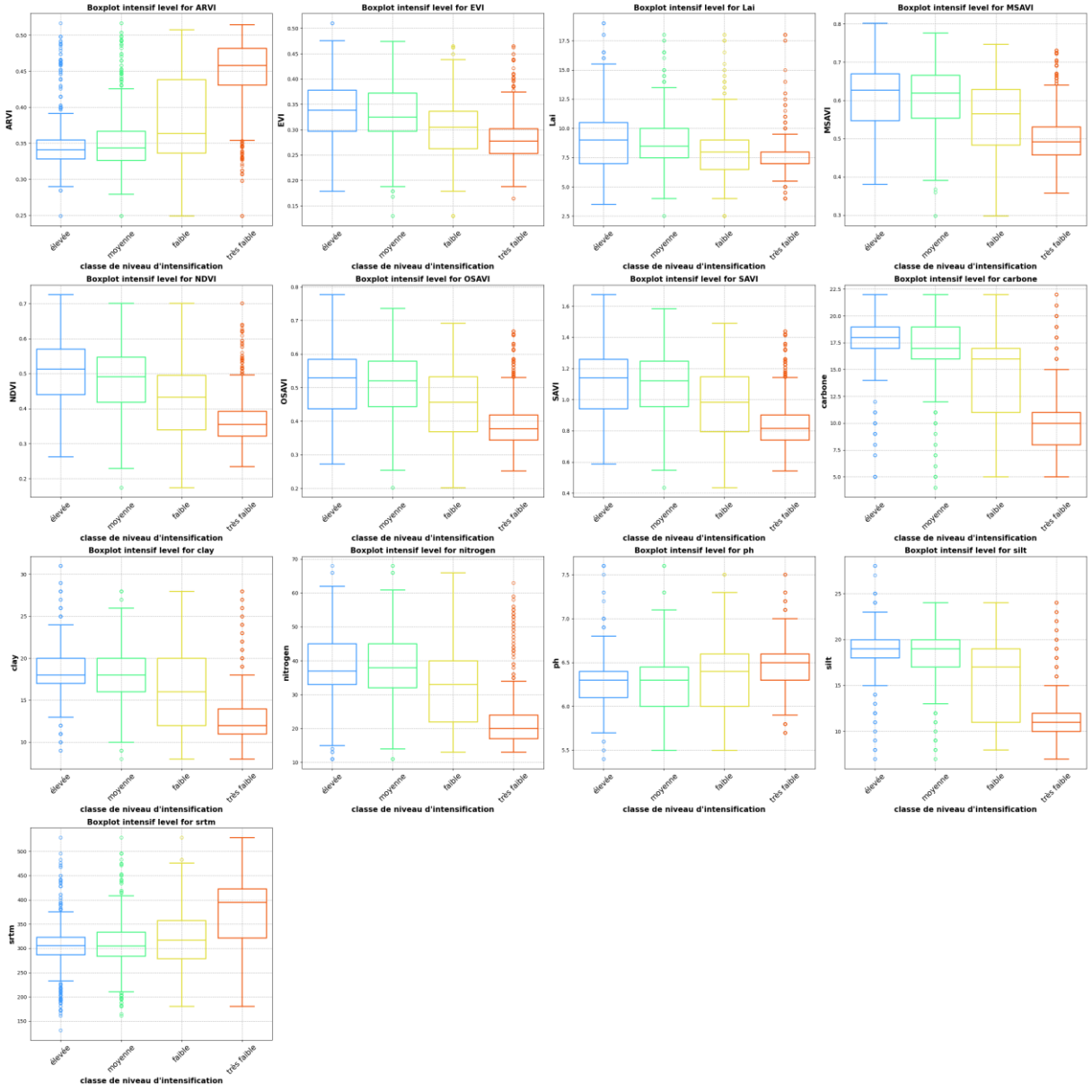


Figure 26: Boxplot des covariables en fonction de chaque classe de niveau d'intensification

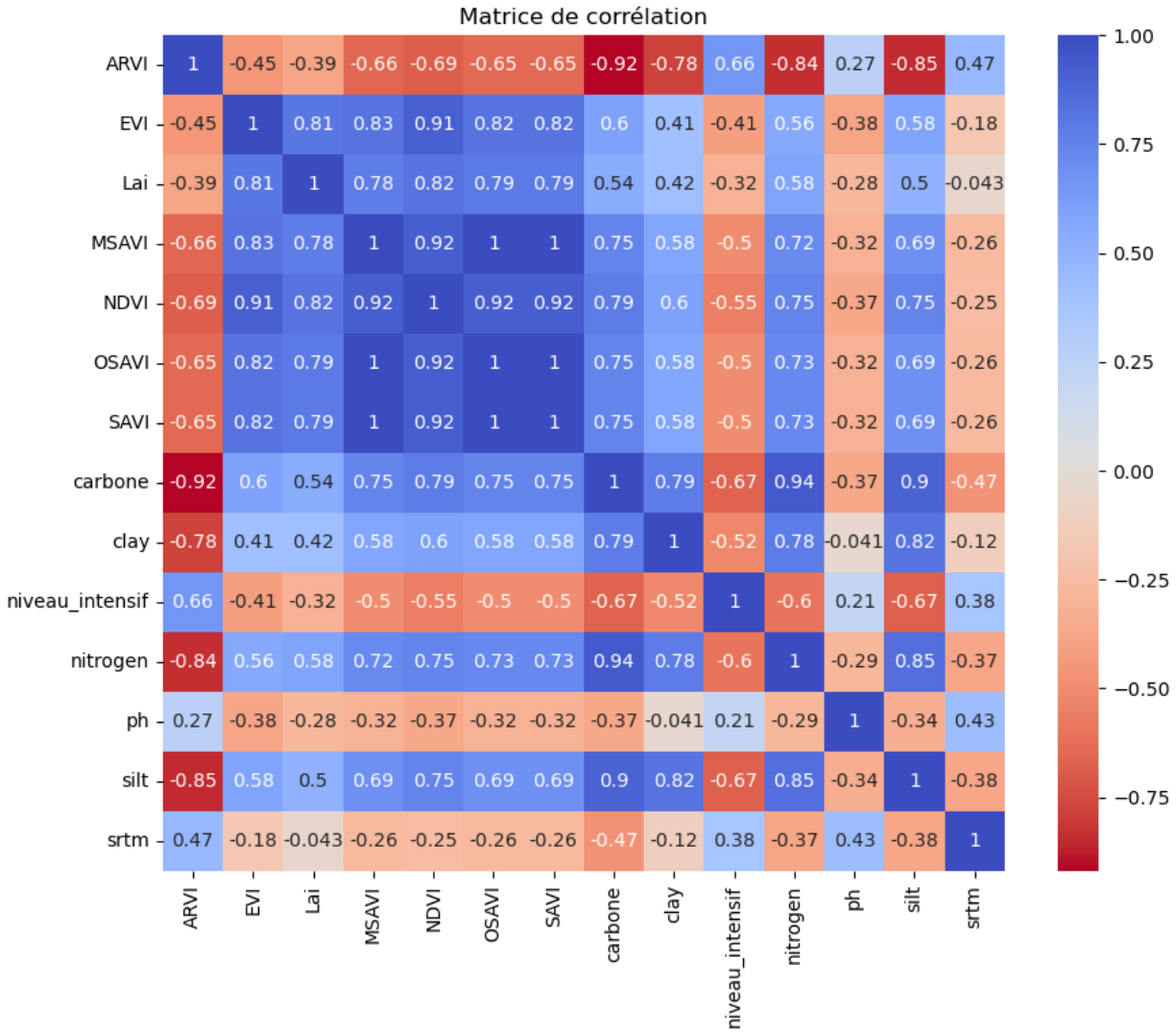


Figure 27: Matrice de corrélation des covariables et du niveau d'intensification

- ***Carte des modèles de régression du yield gap***

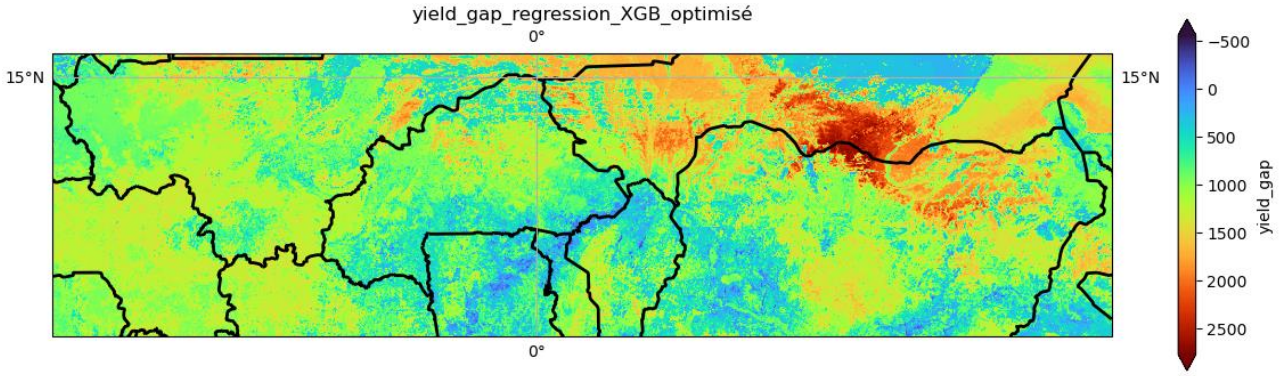


Figure 28: Carte des yield gap avec le XGB optimisé

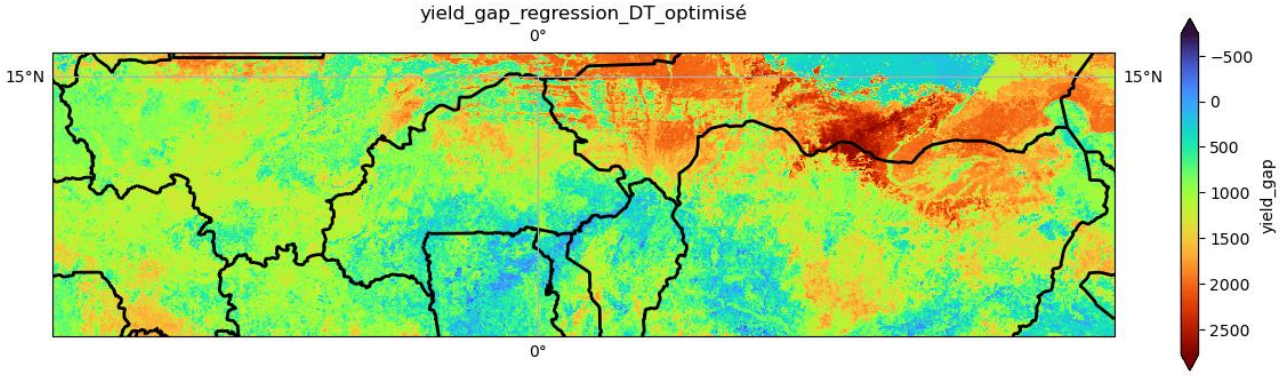


Figure 29: Carte des yield gap avec le DT optimisé

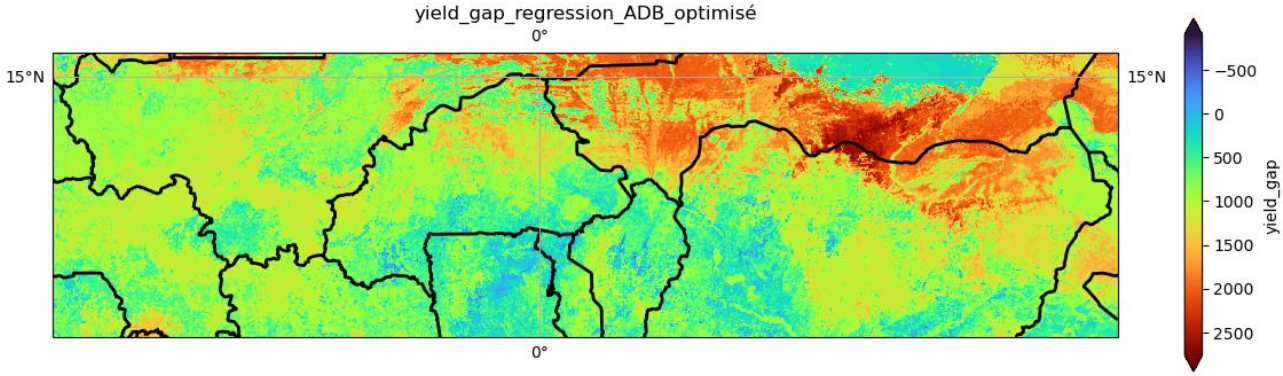


Figure 30: Carte des yield gap avec le ADB optimisé

- ***Carte des modèles de classification du niveau d'intensification optimisés***

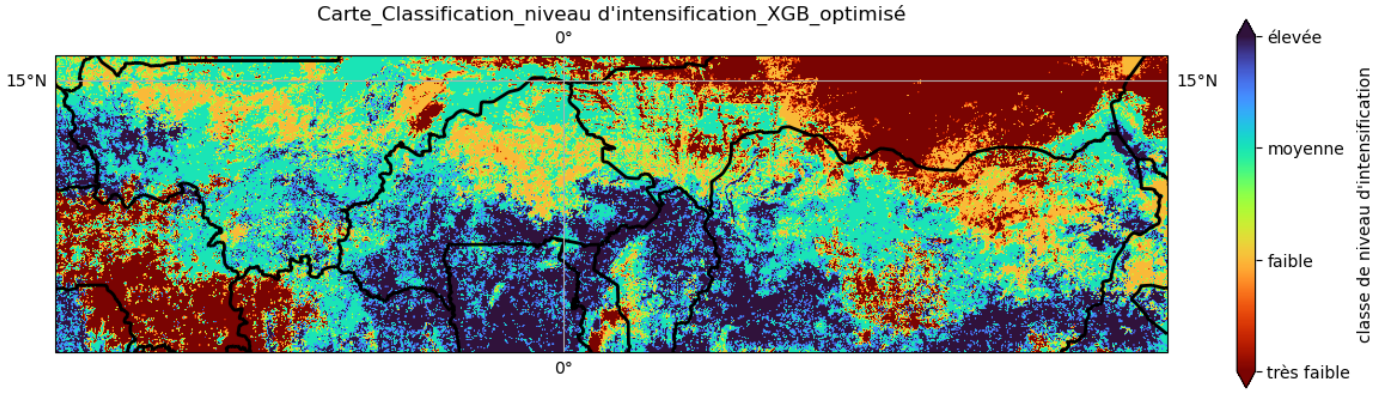


Figure 31: Carte des niveau d'intensification avec le modèle Rf optimisé

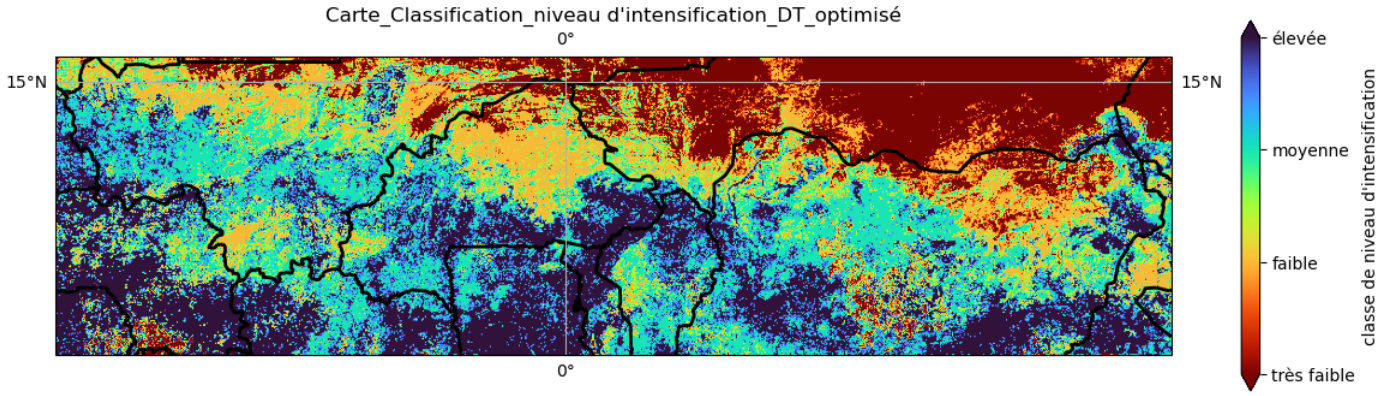


Figure 32: Carte des niveau d'intensification avec le modèle Rf optimisé

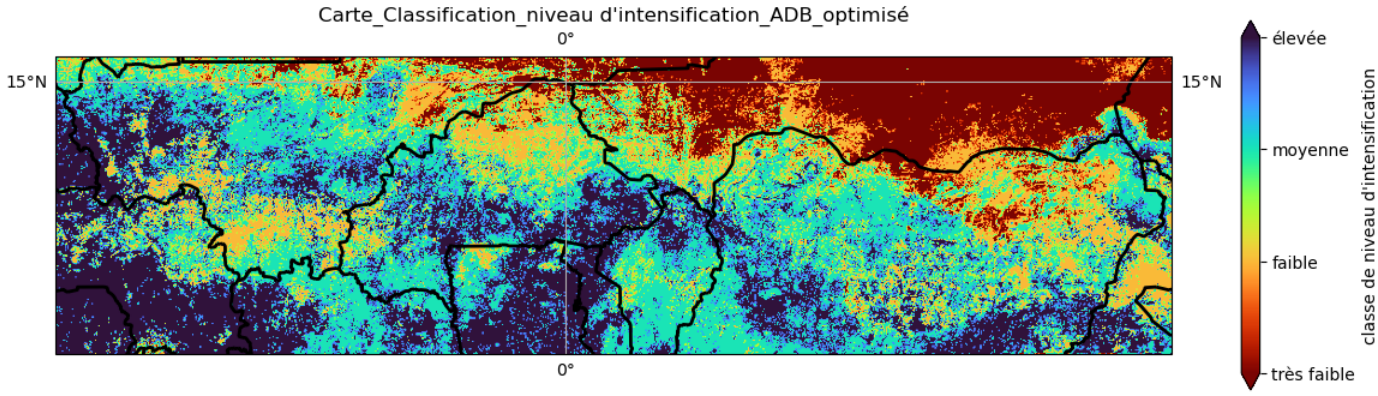


Figure 33: Carte des niveau d'intensification avec le modèle Rf optimisé