



Suivi et Modélisation à l'échelle des pixels de la dynamique de la canne à sucre par télédétection

Pixel-scale Monitoring and Modeling of Sugarcane Dynamics Using Remote Sensing

Ondon Nkoua Cedrick Belmich^{1*}, Samba Lokombe Christian², Tsoala Tchoffo Jessy Houston¹, Okoka Onosamba Jean-Pierre¹, Lundanda Mashindji Rodrigue^{1,4}, Kayengi Baziota Arsene¹, Lokango Okintambolo Olivier², Mandaya Limbeye Felly^{1,3}, Bakili Londa Lexis¹, Isidor Masumbuko Kyamusoke¹ ; Jimy Musitu Berka¹, BAONGI MATELA Josué¹

¹ Apprenant en Master en Géomatique et Changement Climatique, Université de Kinshasa. Faculté des Sciences Agronomiques et Environnement. Département de Gestion des Ressources Naturelles. BP.:117 Kinshasa XI (RDC)

² Centre de Recherches Géologiques et Minières (CRGM), B.P.: 898 Kinshasa/Gombe, R.D. Congo

³ Centre de Recherche en Sciences Humaines (CRESH) B.P.: 3474 Kinshasa/Gombe, R.D. Congo

⁴ Université Pédagogique Nationale, Unité de Recherche Développement des Substances Naturelles et Biodégradables (UPN) B.P.: 8815 Kinshasa/ Binza, R.D. Congo

Abstract: Sugarcane is a strategic crop worldwide, essential for sugar and bioenergy production, and contributes significantly to tropical agricultural economies. In equatorial regions, its productivity depends strongly on climatic and hydrological conditions. Remote sensing provides effective tools to monitor phenological cycles and anticipate yields accurately and reproducibly. This study presents an integrated approach combining Google Earth Engine (GEE) and Python to analyze sugarcane phenology and model productivity. Sentinel-2-time series (2019-2024) and climatic data, including precipitation, temperature, and reference evapotranspiration, were used to calculate multiple vegetation indices, extract key phenological indicators such as the start, peak, and end of season, season duration, and cumulative biomass, detect anomalies, and develop predictive yield models using multiple regression and Random Forest. Results indicate a long cropping season of approximately 354 days, with peak growth from January to March and a slowdown from August to October. Cumulative biomass varies significantly between years and is positively correlated with yield, as well as with the start of season and water-related indices. The Random Forest model outperformed multiple regression, capturing nonlinear interactions and improving predictive accuracy. This automated GEE-Python methodology demonstrates its potential to link phenology, climate, and yield, providing actionable insights for precision agriculture in tropical environments. It also offers a scalable and reproducible framework for monitoring crop development, supporting management decisions, and promoting sustainable intensification of sugarcane production.

Keywords: remote sensing, vegetation indices, phenology, Google Earth Engine, Python, Random Forest, sugarcane, crop productivity

Résumé : La canne à sucre est une culture stratégique à l'échelle mondiale, essentielle à la production de sucre et bioénergie. En zone équatoriale, sa productivité dépend des conditions climatiques et hydriques. La télédétection offre des outils puissants pour suivre les cycles phénologiques et anticiper les rendements. Cette étude propose une approche intégrée combinant Google Earth Engine (GEE) et Python pour analyser la phénologie de la canne à sucre et modéliser sa productivité. Des séries temporelles Sentinel-2 (2019-2024) et

des données climatiques, incluant précipitations, température et évapotranspiration de référence, ont été exploitées pour calculer plusieurs indices de végétation, extraire les indicateurs phénologiques clés tels que le début, le pic et la fin de saison, la durée et la biomasse cumulée, détecter les anomalies et construire des modèles prédictifs de rendement utilisant la régression multiple et Random Forest. Les résultats montrent une saison culturelle longue d'environ 354 jours, avec un pic de vigueur de janvier à mars et un ralentissement d'août à octobre. La biomasse cumulée varie fortement entre années et se corrèle positivement avec le rendement, ainsi qu'avec le début de saison et les indices liés à l'eau. Le modèle Random Forest surpassé la régression multiple, capturant les interactions non linéaires et améliorant la précision prédictive. Cette méthodologie automatisée illustre le potentiel des chaînes GEE-Python pour relier phénologie, climat et rendement, ouvre des perspectives concrètes pour l'agriculture de précision en milieu tropical.

Mots-clés : télédétection, indices de végétation, phénologie, Google Earth Engine, Python, Random Forest, canne à sucre, productivité agricole

Digital Object Identifier (DOI): <https://doi.org/10.5281/zenodo.17839026>

1. INTRODUCTION

L'agriculture demeure un pilier stratégique du développement économique et social à l'échelle mondiale. Parmi les cultures industrielles, la canne à sucre occupe une place de premier plan grâce à sa double contribution à la sécurité alimentaire et énergétique, via la production de sucre et de bioéthanol, tout en générant des emplois et des recettes fiscales significatives [2,1]. Cependant, la gestion de la canne à sucre est confrontée à de nombreux défis : variabilité climatique, dégradation des sols, pression démographique, et limites des méthodes traditionnelles de suivi. L'évaluation du rendement et du développement phénologique repose encore largement sur des observations de terrain, coûteuses et spatialement limitées [3,4]. Ces contraintes freinent la réactivité et l'efficacité des décisions agricoles.

Dans ce contexte, la télédétection spatiale offre une alternative innovante et performante. Grâce aux images satellites à haute résolution, il est désormais possible de suivre la croissance des cultures à l'échelle des pixels, d'extraire des indices de végétation (NDVI, EVI, SAVI), et de modéliser la dynamique phénologique de manière continue [5,7]. L'intégration de ces données avec des variables climatiques permet d'améliorer la prévision des rendements et le suivi des phases clés du cycle culturel [6].

Ainsi, le suivi et la modélisation à l'échelle des pixels de la canne à sucre dans la zone de Nkayi constituent une contribution scientifique et opérationnelle majeure. Cette étude vise à développer une approche intégrée combinant séries temporelles satellitaires et données environnementales pour : identifier les indicateurs phénologiques clés, analyser leur relation avec les variables climatiques, modéliser les rendements agricoles, et produire des visualisations cartographiques utiles à la prise de décision.

Les résultats attendus permettront de renforcer la compréhension des interactions entre paramètres spectraux et agro-écologiques, tout en fournissant des outils d'aide à la décision pour les gestionnaires agricoles et les acteurs de la filière sucrière au Congo[9]. Pour cela nous nous sommes formulés les questions de recherches suivantes :

- Quels sont les principaux indicateurs phénologiques de la canne à sucre détectables par télédétection à l'échelle des pixels dans la région étudiée ?
- Comment les variables climatiques (précipitations, température, évapotranspiration, NDWI) influencent-elles les dynamiques phénologiques de la canne à sucre ?
- Dans quelle mesure les indicateurs phénologiques et climatiques permettent-ils de prédire les rendements de la canne à sucre ?
- Comment la visualisation spatiale et temporelle des données satellitaires peut-elle améliorer la prise de décision agricole à l'échelle locale et régionale ?

Afin de répondre aux questions de recherche posées, nous formulons les hypothèses suivantes :

- Les séries temporelles satellitaires permettent d'identifier avec précision les phases phénologiques clés de la canne à sucre (SOS, POS, EOS, etc.) à l'échelle des pixels.
- Il existe une corrélation significative entre les indicateurs phénologiques et les variables climatiques, influençant directement la croissance et la productivité de la canne à sucre.

- Un modèle prédictif basé sur les données satellitaires et climatiques peut estimer les rendements agricoles avec une précision suffisante pour appuyer la planification.
- La cartographie dynamique des indicateurs NDVI et phénologiques améliore la compréhension spatiale des performances agricoles et facilite la prise de décision.

Cette étude vise à développer une approche méthodique, robuste et opérationnelle pour le suivi et la prédition de la dynamique de la canne à sucre à l'aide de données satellitaires, dans le but de renforcer la planification agricole et la sécurité alimentaire à différentes échelles territoriales.

Les objectifs spécifiques de cette étude se déclinent comme suit :

- Observation et suivi phénologique : identifier les indicateurs phénologiques clés (SOS, POS, EOS, Duration, AUC_NDVI) pour les parcelles de canne à sucre dans la zone d'étude, à partir de séries temporelles satellitaires mensuelles couvrant une période de cinq ans.
- Corrélation avec les variables environnementales : évaluer les relations quantitatives entre les indicateurs phénologiques et les variables climatiques telles que les précipitations, les températures, l'évapotranspiration et l'indice NDWI.
- Modélisation prédictive des rendements : développer un modèle de prédition des rendements basé sur les indicateurs phénologiques et climatiques, calibré à partir de données locales de production.
- Visualisation et diffusion des résultats : produire des cartes mensuelles de NDVI et des cartes annuelles des dates clés (SOS, POS, EOS), accompagnées de visualisations statistiques interactives, afin de faciliter la prise de décision pour les acteurs agricoles et institutionnels.

2. MATERIELS ET METHODE

2.1. Présentation du milieu

La zone d'étude se situe dans le département de la Bouenza, au sud-ouest de la République du Congo (Congo-Brazzaville), autour de la commune de Nkayi (**Figure 1**) et de la localité de Buenza (coordonnées approximatives : 4°15'S, 13°18'E). Elle est caractérisée par un relief plat à légèrement ondulé, constitué de plaines et de bas plateaux favorables aux activités agricoles et à la végétation tropicale dense.

Le climat est de type tropical équatorial à saisonnalité marquée, avec une saison des pluies allant d'octobre à mai et une saison sèche de juin à septembre. Les précipitations annuelles moyennes varient entre 1 200 et 1 600 mm, et la température annuelle moyenne se situe autour de 24–26 °C. Ces conditions climatiques influencent directement la couverture végétale et les usages agricoles de la région.

La zone est desservie par un réseau de routes nationales et rurales, facilitant la mobilité et l'accès aux villages environnants. Elle est également traversée par plusieurs cours d'eau saisonniers et permanents, contribuant à l'irrigation et aux écosystèmes locaux.

Sur le plan démographique, le département de la Bouenza présente une densité moyenne de 28 habitants/km². L'économie locale repose principalement sur l'agriculture, avec une mosaïque de parcelles de cultures vivrières et industrielles, ainsi que sur l'élevage extensif.

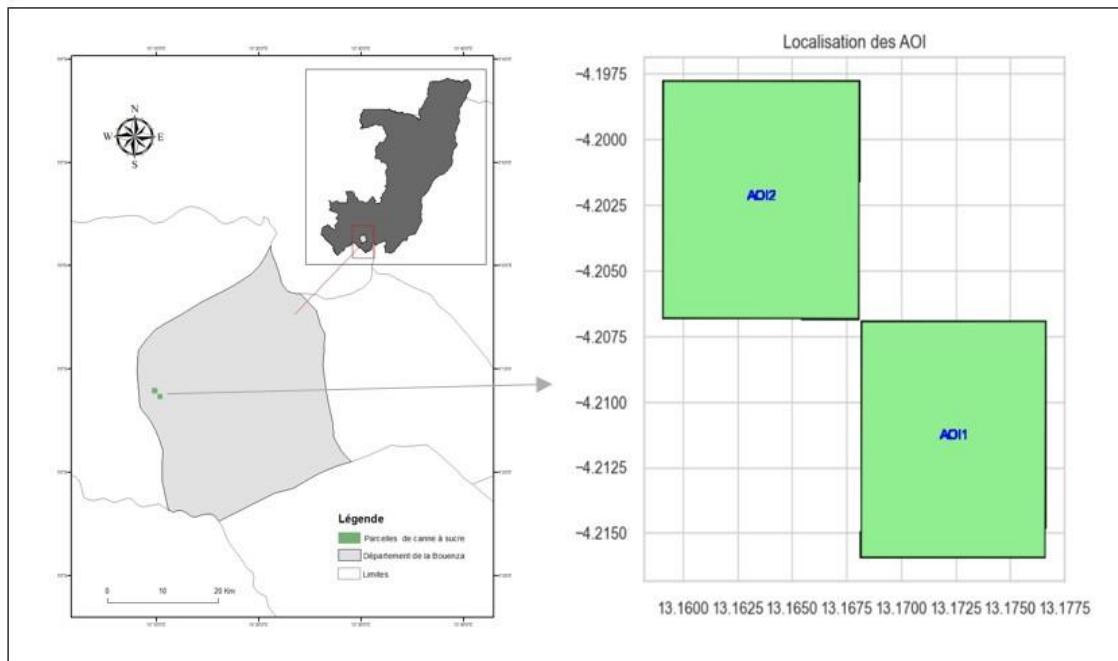


Figure 1. Situations Géographiques de la zone d'étude, les parcelles agricoles définies en AOI

2.2. MATERIELS UTILISES

- Données satellitaires (Sentinel-2) : pour extraire les indices et suivre la croissance.
- Séries temporelles mensuelles sur 5 ans : pour analyser les tendances phénologiques.
- Données climatiques (précipitations, température, évapotranspiration) : pour établir des corrélations avec les indicateurs végétatifs.
- Systèmes SIG (QGIS, ArcGIS, Google Earth Engine) : pour le traitement spatial et la cartographie.
- Outils de modélisation (Python) : pour construire et valider le modèle prédictif.

2.3. METHODOLOGIES

2.3.1. Acquisition des données et prétraitement

La première étape a consisté à développer, dans Google Earth Engine (GEE), un script d'extraction des pixels Sentinel-2 nécessaires au calcul des indices de végétation. Les images Sentinel-2 MSI couvrant la période 2019-2024 ont été extraites via la plateforme Google Earth Engine (GEE), dont l'efficacité pour le traitement massif d'archives satellitaires est largement documentée[11]. Toutes les tuiles intersectant les aires d'intérêt ont été sélectionnées, y compris celles présentant une couverture nuageuse partielle, pratique recommandée pour conserver la continuité des séries en contexte tropical[12].

L'acquisition des données a porté sur l'ensemble des images Sentinel-2 (produits Level-2A) couvrant la période 2019-2024 et intersectant la zone d'étude, située en UTM zone 33 Sud (EPSG :32733). Les bandes spectrales exploitées incluent : Blue (B2, 490 nm, 10 m), Green (B3, 560 nm, 10 m), Red (B4, 665 nm, 10 m), NIR (B8, 842 nm, 10 m), ainsi que SWIR1 (B11, 1610 nm, 20 m) et SWIR2 (B12, 2190 nm, 20 m). Afin de garantir l'homogénéité spatiale, les bandes SWIR à 20 m ont été rééchantillonnées à 10 m (interpolation bilinéaire) et toutes les scènes ont été reprojetées dans le même système de coordonnées (UTM 33S).

Le prétraitement a consisté en : (i) la conversion en réflectance de surface à l'aide de Sen2Cor, (ii) la correction radiométrique par l'élévation solaire selon la relation :

$$R_{corr} = \frac{R_{SR}}{\sin(h)}$$

Où R_{corr} représente la réflectance de surface et h l'angle d'élévation solaire (en radians), et (iii) le masquage des nuages et ombres en appliquant l'algorithme s2cloudless intégré dans GEE, avec un seuil de probabilité fixé à 40 %, seuil qui représente un compromis adapté au contexte tropical de la zone d'étude [13,14].

Un seuil probabiliste de 40 % a été retenu afin de conserver des observations partiellement affectées mais encore exploitables.

En parallèle, les séries climatiques incluant : les précipitations (mm), la température moyenne ($^{\circ}\text{C}$) et l'évapotranspiration de référence (ET0, mm) ont été agrégées, normalisées et intégrées à la base de données pour permettre une analyse conjointe phénologie-climat

2.3.2. Calcul Indices de végétation

Le calcul des indices de végétation permet d'évaluer la vigueur, la biomasse et l'humidité du couvert végétal. Plusieurs indices ont été utilisés : le **NDVI** (Normalized Difference Vegetation Index) [17] qui mesure l'activité photosynthétique ; l'**EVI** (Enhanced Vegetation Index), qui corrige les effets du sol et de l'atmosphère ; le **MSAVI2** (Modified Soil Adjusted Vegetation Index 2) qui réduit l'influence du sol dans les couverts clairsemés ; et enfin l'**NDWI** (Normalized Difference Water Index) [18,19,20], qui évalue l'eau contenue dans la végétation. NDWI (Normalized Difference Water Index, version SWIR).

$$NDVI = \frac{\rho_{NIR} - \rho_{RED}}{\rho_{NIR} + \rho_{RED}}$$

$$EVI = 2.5 \cdot \frac{\rho_{NIR} - \rho_{RED}}{\rho_{NIR} + 6\rho_{RED} - 7.5\rho_{BLUE} + 1}$$

$$MSAVI2 = \frac{2\rho_{NIR} + 1 - \sqrt{(2\rho_{NIR} + 1)^2 - 8(\rho_{NIR} - \rho_{RED})}}{2}$$

$$NDWI_{SWIR} = \frac{\rho_{NIR} - \rho_{SWIR}}{\rho_{NIR} + \rho_{SWIR}}$$

Où ρ_{NIR} et ρ_{RED} représentent respectivement les réflectances NIR (B8) et rouge (B4). Avec $G=2.5$, $C1=6$, $C2=7.5$, $L=1$. Ici, G correspond au facteur de gain, $C1$ et $C2$ aux coefficients de correction atmosphérique (aérosols), et L au facteur d'ajustement du sol. Et ρ_{SWIR} correspond à la bande SWIR1 (B11). Cet indice renseigne sur l'humidité foliaire et l'eau disponible dans la végétation.

Les indices ont été filtrés pour ne retenir que des valeurs biologiquement plausibles ($-0.2 \leq NDVI \leq 0.9$) [21,22] pour les cultures pérennes. Le comblement des lacunes a été réalisé par Whittaker smoothing, qui combine robustesse et préservation des pics de végétation. Ce filtrage permet de préserver la variabilité naturelle tout en éliminant les valeurs issues d'artefacts.

$$X_{filtré} = \{x \in NDVI \mid -0.2 < x < 0.9 \wedge Cloud < 40\%\}$$

2.3.3. Construction du GeoDataFrame

Afin de structurer les données pour les analyses ultérieures, un **GeoDataFrame** (Tableau 1) a été construit. Chaque enregistrement correspond soit à un pixel, soit à une unité spatiale d'intérêt (AOI). Les champs retenus incluent :

Tableau 1. Structure du GéoDataFrame

Nom du champ	Unité	Description
system:index	–	Identifiant unique de l'observation (scène Sentinel-2).
AOI	–	Parcelles agricoles (AOI1, AOI2).
Date	YYYY-MM-DD	Date d'acquisition de l'image Sentinel-2.
CloudFlag	fraction (0–1)	Fraction nuageuse détectée sur la scène.
NDVI	adim.	Normalized Difference Vegetation Index.
EVI	adim.	Enhanced Vegetation Index.
MSAVI2	adim.	Modified Soil Adjusted Vegetation Index 2.

GNDVI	adim.	Green Normalized Difference Vegetation Index.
NDWI	adim.	Normalized Difference Water Index.
SAVI	adim.	Soil Adjusted Vegetation Index.
MoyMens	adim.	Moyennes mensuelles des indices spectraux (NDVI, EVI, etc.).
MoyAnn	adim.	Moyennes annuelles des indices spectraux (NDVI, EVI, etc.).
Year	Année	Année d'observation.
Month, Month_	Mois	Mois de l'acquisition, utile pour l'analyse saisonnière.
rec(mm/m)	mm/mois	Précipitations mensuelles cumulées.
tmp_min(°C)	°C	Température minimale mensuelle.
tmp_max(°C)	°C	Température maximale mensuelle.
tmp_mean(°C)	°C	Température moyenne mensuelle.
rel_Hum(%)	%	Humidité relative moyenne.
Sunshine(J/m ²)	J/m ²	Rayonnement solaire cumulé mensuel.
Wind(2m)(m/s)	m/s	Vitesse moyenne du vent à 2 m.
ETo(mm/m)	mm/mois	Évapotranspiration de référence.
geometry	Polygone	Géométrie représentant la parcelle ou AOI (en EPSG :32733).

Ce GeoDataFrame a été exporté sous format GeoPackage (.gpkg), garantissant une intégration aisée dans les environnements SIG et facilitant les traitements spatio-temporels.

2.3.4. Extraction des indicateurs phénologiques

Les séries temporelles de NDVI ont été utilisées pour dériver les paramètres clés du cycle de croissance : **début de saison (SOS)**, **pic de saison (POS)** et **fin de saison (EOS)** [23], ces indicateurs sont estimés à partir des dérivées temporelles du NDVI :

$$SOS = \min\{t \mid NDVI'(t) > \tau\} \quad POS = \arg \max_t NDVI(t) \quad EOS = \max\{t \mid NDVI'(t) < -\tau\}$$

Où SOS c'est le premier instant t où la dérivée du NDVI (NDVI'(t)) dépasse un seuil positif t, POS c'est l'instant où le NDVI atteint sa valeur maximale, EOS c'est le dernier instant où la dérivée du NDVI devient inférieure à -t. On identifie ainsi quand la végétation décroît fortement (fin de la saison verte). Autrement dit, le moment du « plein développement » de la végétation. On repère donc quand la végétation commence réellement à croître après la saison sèche.

La durée de la saison est calculée par : Duration = EOS-SOS. La durée de la saison correspond simplement au temps écoulé entre le début et la fin de la phase active de végétation.

La biomasse cumulée est estimée par l'aire sous la courbe du NDVI : $AUC_{NDVI} = \sum_{i=1}^{n-1} \frac{NDVI_i + NDVI_{i+1}}{2} (t_{i+1} - t_i)$

On divise la saison en n intervalles (t_i, t_{i+1}), NDVI_i est la valeur de l'indice au temps t_i. On utilise la méthode des trapèzes : moyenne des NDVI sur l'intervalle × durée de l'intervalle. Cela revient à intégrer le NDVI au cours du temps, ce qui donne une estimation de la quantité de biomasse accumulée pendant la saison.

2.3.5. Détection d'anomalies et ruptures temporelles

Afin d'identifier d'éventuels stress climatiques, perturbations environnementales ou erreurs de mesure dans les séries temporelles, une analyse des anomalies a été effectuée à l'aide du **z-score** [24].

$$z_i = \frac{NDVI_i - \overline{NDVI}}{\sigma_{NDVI}}$$

Où \overline{NDVI} est la moyenne de l'ensemble des observations et σ leur écart-type. Cette normalisation permet de mesurer combien une observation s'écarte de la distribution moyenne de la série

Une observation a été considérée comme anomalie si la valeur absolue de son z-score dépassait 2,5 ($|z_i| > 2,5$). Les observations identifiées comme anomalies ont été analysées séparément, afin de distinguer les variations normales des ruptures significatives ou des erreurs potentielles dans les données. Cette approche permet de garantir la qualité des séries temporelles et d'optimiser l'interprétation des indicateurs dérivés des images satellitaires

2.3.6. Corrélations entre variables

Les relations entre indicateurs phénologiques, variables climatiques et rendement ont été quantifiées par le coefficient de corrélation de Pearson :

$$r_{XY} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2 \sum (Y_i - \bar{Y})^2}}$$

Où X_i : valeur de la variable X pour l'observation i et Y_i valeur de la variable Y pour l'observation i. Les restes se sont les moyennes X et Y.

2.3.7. Modélisation prédictive

Les liens entre les variables climatiques, les indicateurs phénologiques et les rendements ont été évalués à l'aide du coefficient de corrélation de Pearson [25]. Sur cette base, un modèle de régression linéaire a été construit pour prédire le rendement :

$$Yield = \beta_0 + \beta_1 SOS + \beta_2 POS + \beta_3 EOS + \beta_4 AUC + \beta_5 NDVI_{moy} + \epsilon$$

Pour compléter cette approche linéaire, un modèle Random Forest a été employé afin de capturer d'éventuelles relations non linéaires et interactions complexes entre les variables [26,28]. La prédiction du rendement pour un jeu de variables X est donnée par :

$$\hat{y}(x) = \frac{1}{M} \sum_{m=1}^M h_m(x)$$

Où $y(x)$: prédiction du rendement pour un jeu de variables x. M : nombre total d'arbres dans la forêt. $h_{m(x)}$: prédiction donnée par l'arbre m pour l'observation x. Le modèle Random Forest agrège les prédictions de plusieurs arbres de décision pour tenir compte de relations non linéaires et interactions complexes.

II.3.8. Évaluation des modèles

La performance des modèles a été évaluée à l'aide du coefficient de détermination (R^2), de la racine de l'erreur quadratique moyenne (RMSE) et de l'erreur absolue moyenne (MAE), conformément aux recommandations de Willmott et Matsuura (2005) :

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \quad RMSE = \sqrt{\frac{1}{n} \sum (y_i - \hat{y}_i)^2} \quad MAE = \frac{1}{n} \sum |y_i - \hat{y}_i|$$

II.3.9. Visualisation

Les résultats ont été représentés sous forme de cartes mensuelles de NDVI et de cartes annuelles de SOS, POS et EOS, générées dans Python puis affinées dans un SIG. Ces produits visuels facilitent la compréhension spatiale des dynamiques et complètent l'interprétation statistique.

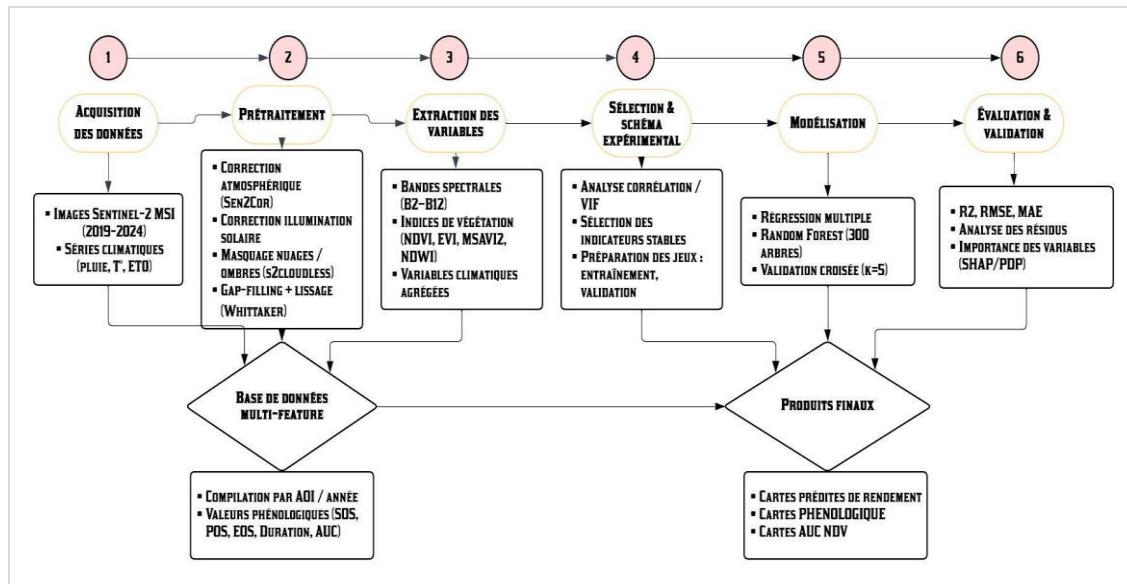


Figure 2. Schéma simplifié décrivant les principales méthodes pour notre étude

3. RESULTATS ET DISCUSSION

3.1. Présentations des résultats

3.1.1. Qualité et couverture des données satellitaires

L'analyse des images Sentinel-2 acquises entre 2019 et 2024 a permis de constituer une base de données robuste pour le suivi de la canne à sucre. Après correction atmosphérique et masquage des nuages via l'algorithme *s2cloudless* (seuil fixé à 40 %), un total de 864 scènes a été retenu. Le nombre d'images exploitables varie selon les années, reflétant l'influence de la couverture nuageuse : la proportion la plus faible a été observée en 2023 (17,4 %), tandis que l'année 2024 présente la couverture la plus élevée (29,4 %). La moyenne sur l'ensemble de la période est de 23,8 %. Cette distribution interannuelle est présentée dans le **Tableau 2** et illustrée à la **Figure 3**, qui met en évidence les fluctuations liées au régime climatique local.

Tableau 2. Répartition annuelle des images Sentinel-2 retenues (2019–2024)

Année	Total scènes	% scènes retenues
2019	140	22,86
2020	144	21,53
2021	144	28,47
2022	146	23,29
2023	144	17,36
2024	146	29,45

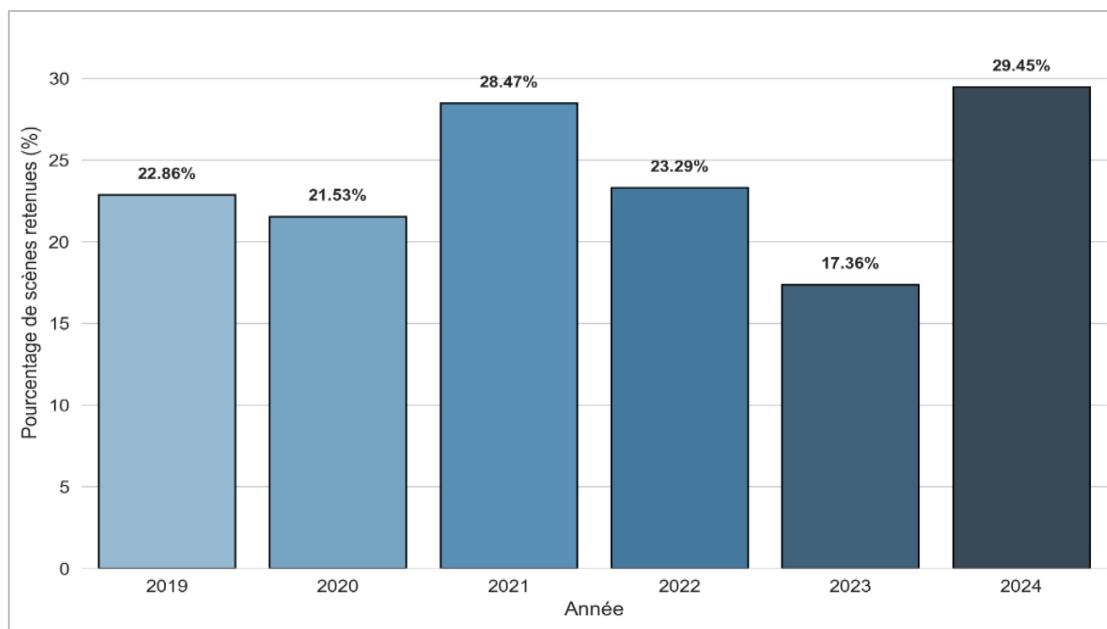


Figure 3. Proportion annuelle des scènes Sentinel-2 retenues après prétraitement et masquage des nuages (2019–2024).

3.1.2. Dynamique saisonnière des indices de végétation

3.1.2.1. Statistiques descriptives globales

Les statistiques descriptives calculées sur l'ensemble de la période montrent des différences notables entre indices spectraux (Tableau 3). Le NDVI, indicateur de vigueur et de biomasse, varie de $-0,090$ à $0,765$ avec une moyenne de $0,136 \pm 0,173$ et un maximum observé en mars. L'EVI, conçu pour réduire les effets du sol et de l'atmosphère, présente une amplitude plus marquée ($-9,697$ à $12,100$; moyenne $0,281 \pm 1,585$), avec un maximum en janvier. Le MSAVI2, adapté aux zones clairsemées, atteint des valeurs comprises entre $-0,198$ et $0,866$ (moyenne $0,202 \pm 0,226$), avec un pic également en mars. Enfin, le NDWI, sensible aux variations en eau du couvert, présente des valeurs globalement négatives ($-0,660$ à $0,125$; moyenne $-0,125 \pm 0,161$), avec un maximum en mai.

Tableau 3. Statistiques descriptives globales des indices de végétation (2019–2024)

Indice	AOI	count	mean	std	min	25%	50%	75%	max
NDVI	AOI1	432	0.139	0.174	-0.090	0.022	0.069	0.213	0.746
EVI	AOI1	432	0.300	1.594	-9.697	-0.324	0.025	0.713	6.765
NDWI	AOI1	432	-0.127	0.162	-0.649	-0.222	-0.066	-0.017	0.125
SAVI	AOI1	432	0.208	0.261	-0.134	0.033	0.103	0.319	1.119
NDVI	AOI2	432	0.134	0.172	-0.090	0.019	0.072	0.180	0.765
EVI	AOI2	432	0.262	1.576	-6.278	-0.391	0.017	0.536	12.100
NDWI	AOI2	432	-0.123	0.161	-0.660	-0.199	-0.066	-0.013	0.124
SAVI	AOI2	432	0.201	0.258	-0.135	0.029	0.107	0.270	1.147

3.1.2.2. Variations mensuelles et interannuelles

Les moyennes mensuelles confirment une dynamique saisonnière nette. Le NDVI et le SAVI atteignent leurs maxima entre janvier et mars, traduisant une forte activité photosynthétique en début de cycle, avant de diminuer progressivement vers octobre où les minima sont atteints. L'EVI conserve une structure saisonnière similaire mais montre une variabilité interannuelle plus importante, ce qui traduit des différences de conditions de croissance entre années. Le NDWI suit une évolution contrastée, avec des valeurs faibles au début de l'année (déficit hydrique relatif) et une amélioration en fin de cycle. Ces dynamiques sont illustrées dans les [Figure 4](#) et les [Figure 5](#).

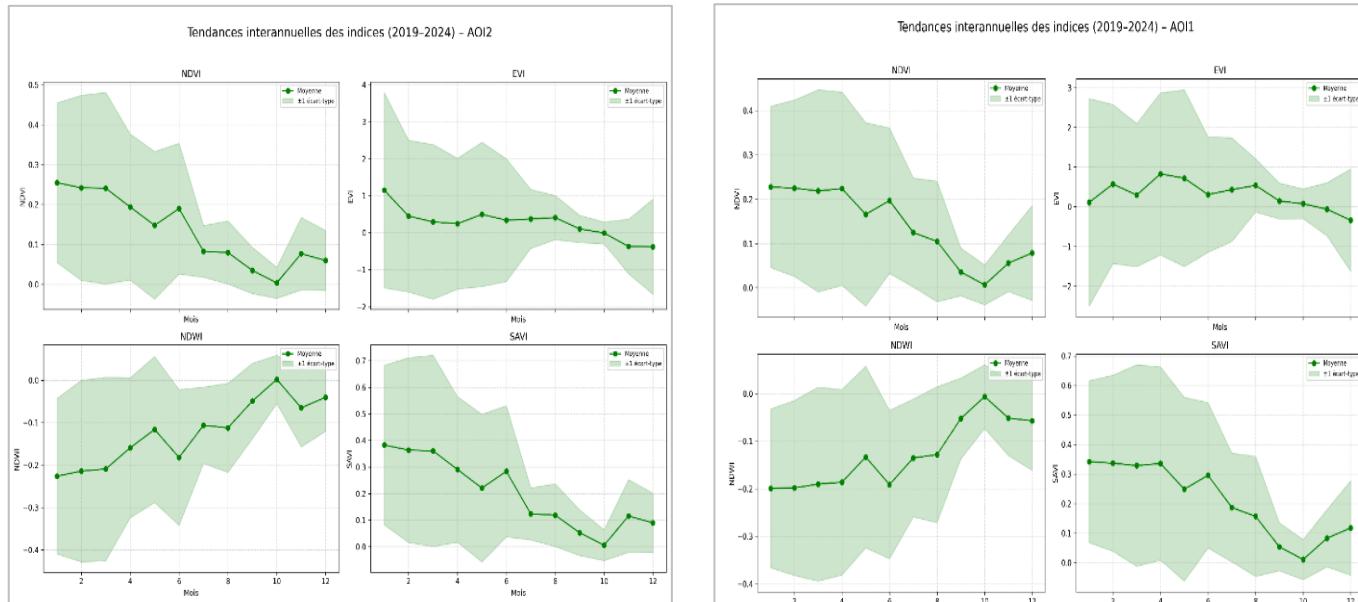


Figure 4. Les tendances interannuelles des indices (moyenne \pm 1 écart-type) pour AOI1 et AOI2 (2019–2024).

3.1.3. Dynamiques saisonnières interannuelles

Les courbes mensuelles permettent d'observer la variabilité interannuelle :

- Les maxima des indices apparaissent en début d'année (janvier–mars).
- Les minima sont localisés entre août et octobre.
- Les profils sont similaires pour les deux AOI, avec des valeurs légèrement plus élevées pour AOI2 sur certains indices (notamment NDVI et EVI).

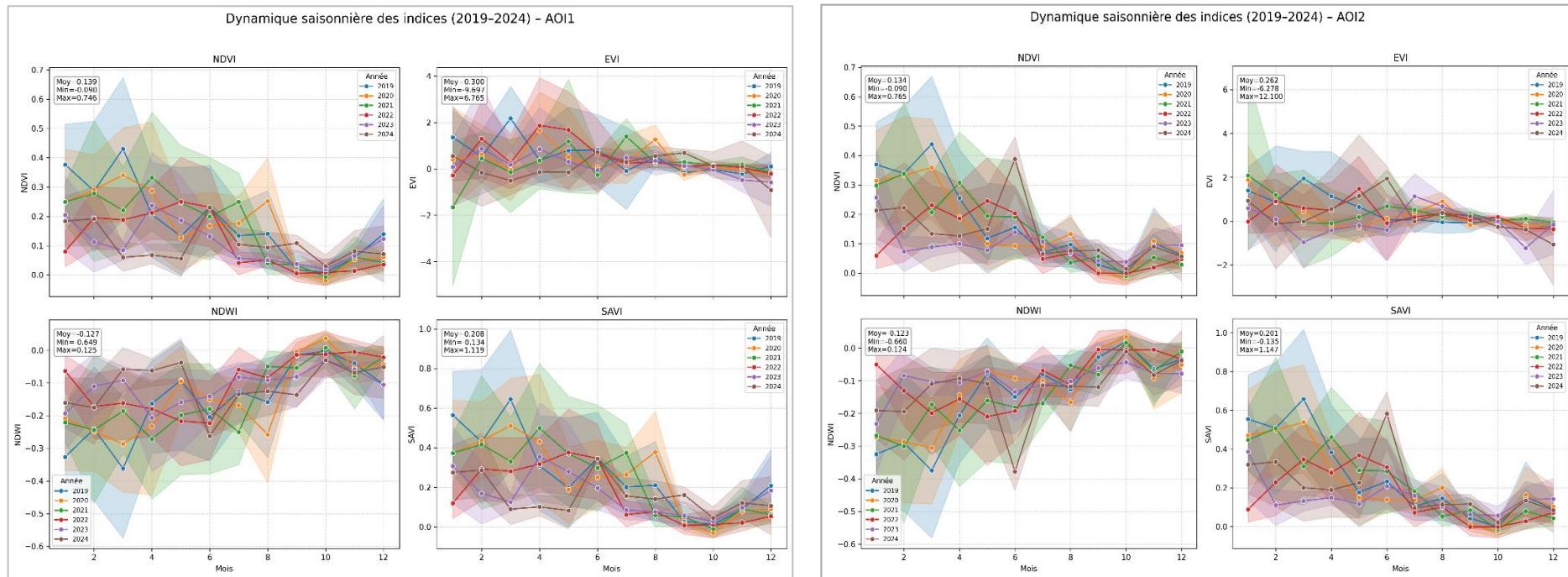


Figure 5. Dynamique saisonnière interannuelle des indices NDVI, EVI, NDWI et SAVI (2019–2024)

3.1.3. Relations entre indices et climat

L'analyse corrélative met en évidence des liens significatifs entre les indices spectraux et les variables climatiques (**Tableau 4** et **Figure 6**). Le NDVI est positivement corrélé à la température moyenne ($r = 0,161$; $p < 0,001$), à l'évapotranspiration de référence ($r = 0,206$; $p < 0,001$), à l'humidité relative ($r = 0,179$; $p < 0,001$) et à l'ensoleillement ($r = 0,290$; $p < 0,001$). Une corrélation négative est observée avec la vitesse du vent ($r = -0,235$; $p < 0,001$). Le MSAVI2 présente un comportement similaire, tandis que le NDWI est corrélé négativement avec l'ET₀, l'humidité et l'ensoleillement, mais positivement au vent. L'EVI ne présente pas de corrélations significatives, ce qui souligne son indépendance relative vis-à-vis des conditions climatiques locales.

Tableau 4. Corrélations de Pearson (r) et p-values entre indices de végétation et variables

Extrait — corrélations (r)

Indice	rec(mm/m)	tmp_mean(°C)	ETo(mm/m)	rel_Hum(%)	Sunshine	Wind
NDVI	0.013	0.161	0.206	0.179	0.290	-0.235
EVI	-0.090	-0.004	-0.004	-0.055	0.019	0.009
MSAVI2	-0.003	0.137	0.186	0.175	0.277	-0.231
NDWI	0.063	-0.063	-0.121	-0.093	-0.202	0.137

Extrait — p-values

Indice	rec(mm/m)	tmp_mean(°C)	ETo(mm/m)	rel_Hum(%)	Sunshine	Wind
NDVI	0.7029	0.0000	0.0000	0.0000	0.0000	0.0000
EVI	0.0078	0.8993	0.9026	0.1062	0.5725	0.7838
MSAVI2	0.9406	0.0001	0.0000	0.0000	0.0000	0.0000
NDWI	0.0639	0.0657	0.0004	0.0065	0.0000	0.0001

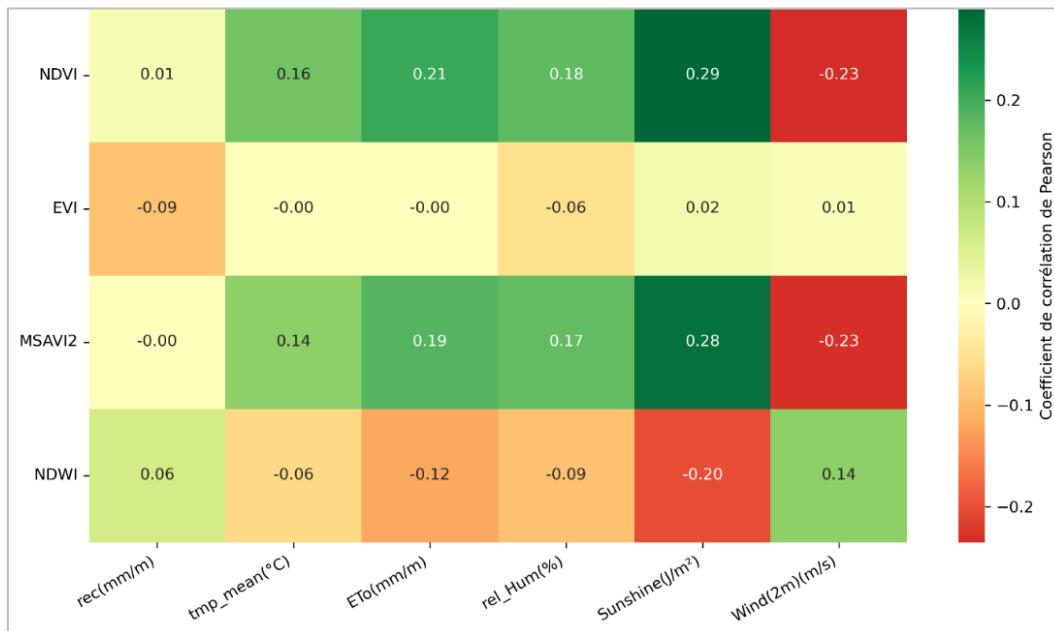


Figure 6. Corrélations entre indices de végétation et variables climatiques (2019–2024)

3.1.4. Indicateurs phénologiques

Les indicateurs phénologiques dérivés du NDVI (SOS, POS, EOS, durée de la saison et AUC-NDVI) mettent en évidence une grande stabilité du cycle cultural de la canne à sucre (**Tableau 5**) (**Tableau 6**). Dans AOI1, la durée moyenne de la saison est de 354 jours, avec un SOS moyen fixé au jour julien 5, un POS au jour 94 et un EOS au jour 360. L'AUC-NDVI moyen est de 9,84, avec des valeurs extrêmes allant de 12,82 (2019) à 7,32 (2024). Dans AOI2, la durée moyenne atteint 355 jours, le POS est légèrement décalé (jour 99), et l'AUC-NDVI moyen est de 9,47, avec un maximum de 12,06 en 2019 et un minimum de 6,96 en 2023. Ces résultats sont représentés dans la **Figure 7**, qui illustre la variabilité interannuelle des stades phénologiques et de la biomasse cumulée.

Tableau 5. Phénologie NDVI pour AOI1 et AOI2 (2019–2024)

AOI	Year	SOS	POS	EOS	AUC-NDVI
AOI1	2019	2019-01-06	2019-03-02	2019-12-27	12.820
AOI1	2020	2020-01-06	2020-04-05	2020-12-26	11.778
AOI1	2021	2021-01-05	2021-04-20	2021-12-26	11.404
AOI1	2022	2022-01-05	2022-05-25	2022-12-21	7.892
AOI1	2023	2023-01-05	2023-05-05	2023-12-21	7.833
AOI1	2024	2024-01-05	2024-01-30	2024-12-30	7.321
AOI2	2019	2019-01-06	2019-03-02	2019-12-27	12.058
AOI2	2020	2020-01-06	2020-03-31	2020-12-26	10.487
AOI2	2021	2021-01-05	2021-04-20	2021-12-26	10.509
AOI2	2022	2022-01-05	2022-05-25	2022-12-21	7.771
AOI2	2023	2023-01-05	2023-01-20	2023-12-26	6.956
AOI2	2024	2024-01-05	2024-06-13	2024-12-30	9.060

Tableau 6. Statistiques phénologiques résumées (AOI1 / AOI2)

Indicateur	AOI1	AOI2
Durée moyenne de la saison (jours)	354.2	355.0
AUC-NDVI moyen	9.84	9.47
SOS moyen (jour julien)	5	5
POS moyen (jour julien)	94	99
EOS moyen (jour julien)	360	360
AUC maximal (année)	12.82 (2019)	12.06 (2019)
AUC minimal (année)	7.32 (2024)	6.96 (2023)

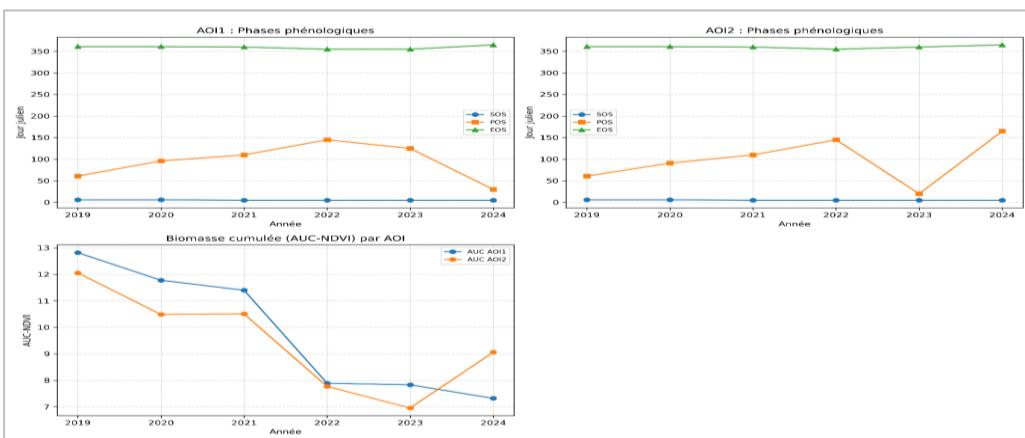


Figure 7. Synthèse des indicateurs phénologiques NDVI (AOI1 et AOI2)

3.1.5. Détection d'anomalies NDVI

L'application d'un score z ($|z| > 2,5$) a permis d'identifier 12 anomalies significatives sur l'ensemble de la période, dont 8 dans AOI1 et 4 dans AOI2. Ces anomalies concernent principalement les mois de mai, juin, août, octobre et décembre. Dans AOI1, les NDVI extrêmes associés varient entre 0,130 et 0,708, tandis que dans AOI2 ils oscillent entre 0,105 et 0,622. Ces événements atypiques, présentés dans le [Tableau 7](#), sont représentés graphiquement dans la [Figure 8](#) (distribution temporelle des anomalies).

Tableau 7. Anomalies NDVI par AOI

AOI	Nombre total d'anomalies	Année(s) concernée(s)	Mois(s) concerné(s)	NDVI min	NDVI max
AOI1	8	2019, 2020, 2021, 2023, 2024	5, 6, 8, 10, 11, 12	0.130	0.708
AOI2	4	2019, 2020, 2021, 2023	5, 10, 11	0.105	0.622

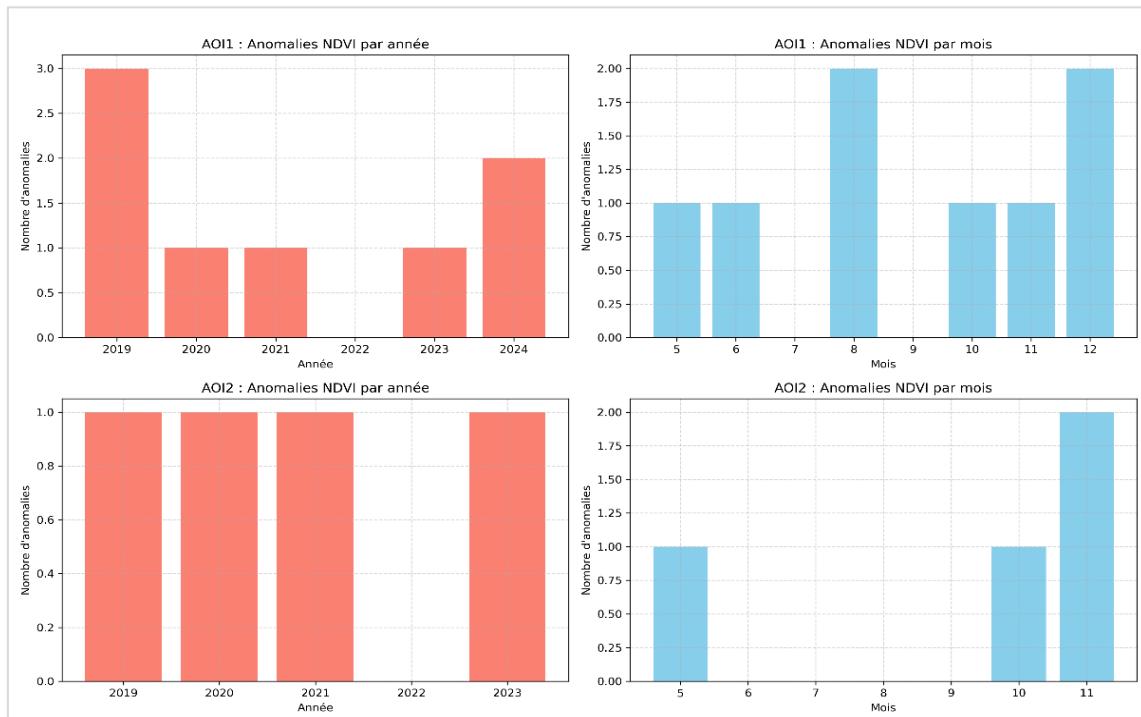


Figure 8. Distribution des anomalies NDVI par AOI

3.1.6 Corrélation et modélisation du rendement

3.1.6.1 Corrélation

L'analyse statistique met en évidence des relations claires entre certaines variables spectrales ou phénologiques et le rendement observé [Tableau 8](#). Les plus fortes corrélations sont observées pour le SOS ($r = 0,390$), l'AUC-NDVI ($r = 0,384$) et le NDWI moyen ($r = 0,342$). Le NDVI moyen et les précipitations annuelles présentent également des corrélations positives mais plus modérées ($r = 0,327$ et $r = 0,247$ respectivement).

Tableau 8. Corrélations de Pearson entre le rendement et les variables explicatives

Variable	Corrélation avec Rendement
SOS (jour julien)	0.390
AUC-NDVI	0.384
NDWI moyen	0.342
NDVI moyen	0.327
Rec mm/an	0.247
ETo annuel	0.130
POS	0.062
EOS (jour julien)	0.042
Température moyenne annuelle	0.025

3.1.6.2. Modélisation et validation

Deux approches de modélisation ont été testées pour la prédiction du rendement : une régression linéaire multiple et un modèle Random Forest.

Les coefficients estimés pour la régression linéaire ([Tableau 9](#)) montrent que le SOS (+0,742) et l'AUC-NDVI (+10,415) contribuent positivement au rendement, tandis que le POS (-6,846), le NDVI moyen (-2,448), le NDWI moyen (-0,409) et l'EOS (-0,415) présentent des coefficients négatifs.

Dans le modèle Random Forest, les variables les plus influentes sont le SOS (importance = 0,291) et le POS (0,238), suivis du NDWI moyen (0,146), du NDVI moyen (0,126) et de l'AUC-NDVI (0,113). L'EOS apparaît comme la variable ayant la plus faible contribution (0,086) ([Tableau 10](#)).

La comparaison des performances des deux modèles ([Tableau 11](#)) met en évidence un meilleur ajustement du Random Forest, avec un R^2 de 0,818 et un RMSE de 2,534, contre un R^2 de 0,587 et un RMSE de 3,816 pour la régression linéaire. Le coefficient de corrélation entre rendements observés et prédis est de 0,962 pour le Random Forest et de 0,766 pour la régression linéaire.

L'équation finale de la régression multiple, construite à partir des variables standardisées, est la suivante :

$$\text{Yield} = 53.7942 + 0.7424 * \text{SOS} + -0.4148 * \text{EOS} + -6.8457 * \text{POS} + -2.4481 * \text{NDVI} + 10.4152 * \text{AUC_NDVI} + -0.4089 * \text{NDWI}$$

Les graphiques associés présentent la calibration des modèles ([Figure 9](#)), la dépendance partielle des variables dans le modèle Random Forest ([Figure 10](#)) ainsi que la relation entre rendements observés et prédis ([Figure 11](#)).

Tableau 9. Coefficients standardisés de la régression linéaire multiple

Variable	Coefficient
SOS	0.742
EOS	-0.415
POS	-6.846
NDVI_moy	-2.448
AUC_NDVI	10.415
NDWI_moy	-0.409

Tableau 10. Importances des variables dans le modèle Random Forest

Variable	Importance RF
SOS	0.291
POS	0.238
NDWI_moy	0.146
NDVI_moy	0.126
AUC_NDVI	0.113
EOS	0.086

Tableau 11. Métriques des modèles

Modèle	R ²	RMSE	MAE
Régression linéaire	0.587	3.816	3.036
Random Forest	0.818	2.534	2.331

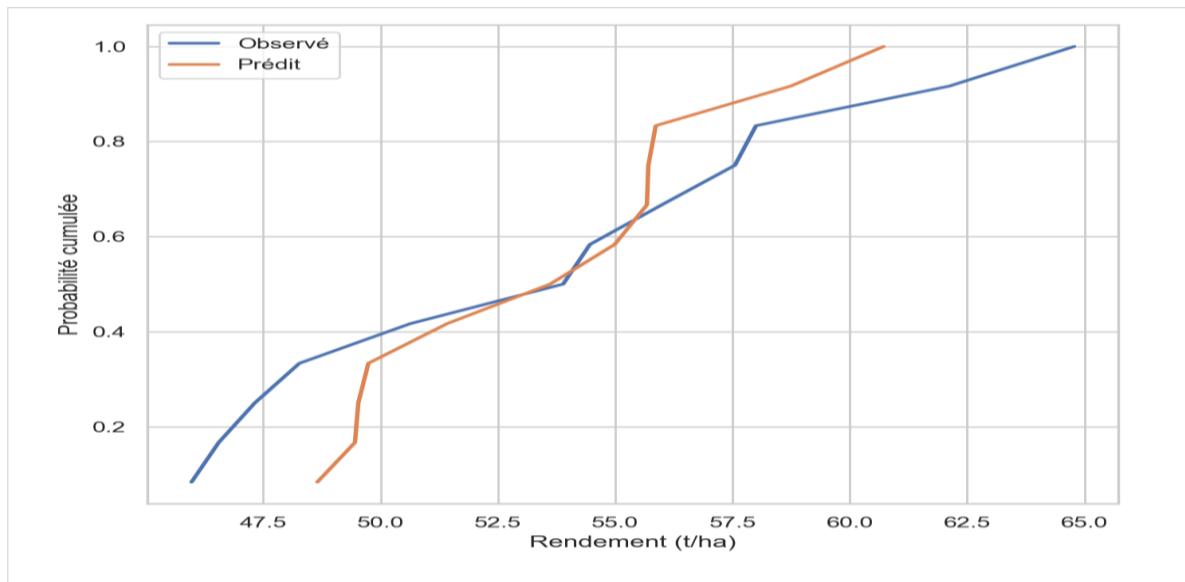


Figure 9. Calibration cumulée

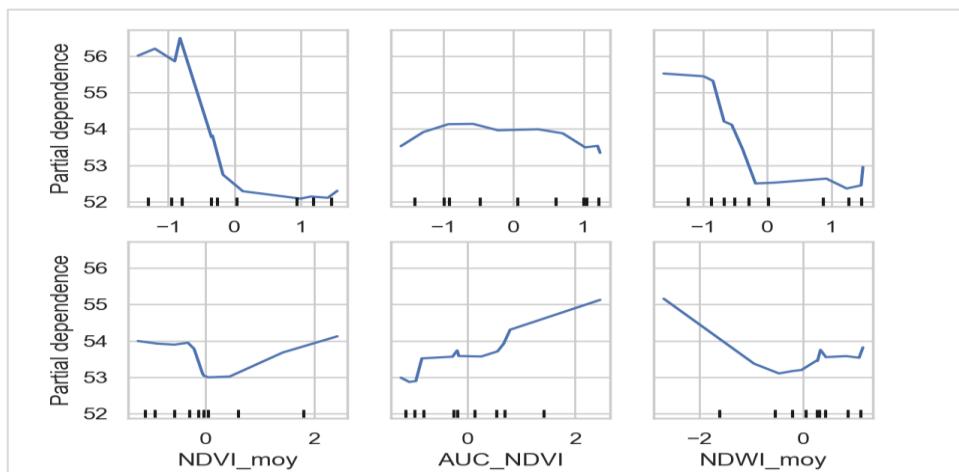


Figure 10. La dépendance partielle des variables dans le modèle Random Forest

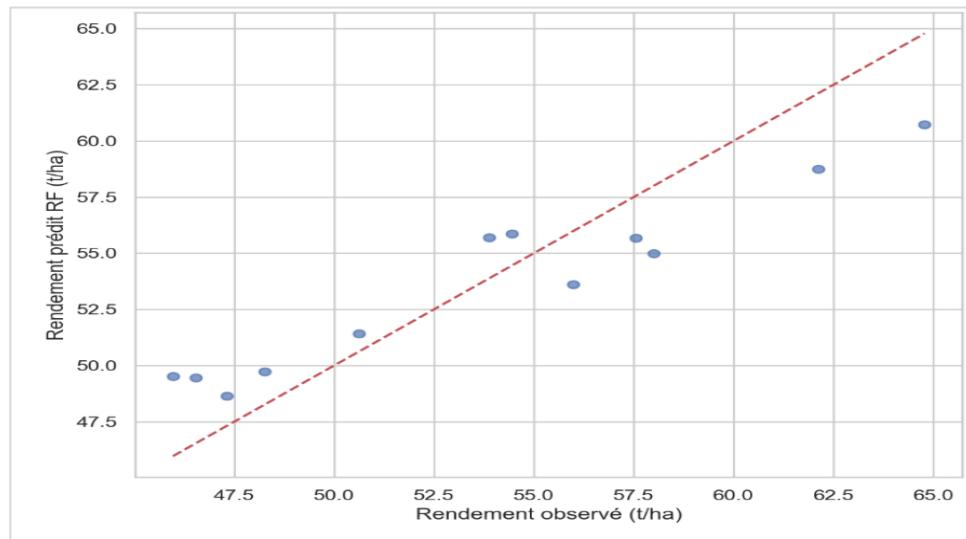


Figure 11. Relation entre rendement observé et prédit

3.1.7. Visualisation cartographique

La dimension spatiale de l'étude a été analysée à travers la production de plusieurs cartes thématiques issues des séries Sentinel-2. Les cartes phénologiques, représentant le début de saison (SOS), le pic de saison (POS) et la fin de saison (EOS), sont présentées dans la Figure 12. Elles mettent en évidence l'hétérogénéité spatiale du développement de la canne à sucre au sein des zones étudiées, en soulignant des variations locales dans la synchronisation et la durée des phases phénologiques.

L'aire sous la courbe du NDVI (AUC-NDVI), indicateur de la biomasse cumulée, a également été cartographiée pour chaque campagne (Figure 13). Ces cartes permettent d'identifier les secteurs caractérisés par un fort potentiel productif et, inversement, de localiser les zones où la croissance est plus limitée.

Enfin, les cartes de rendement prédict (Figure 14) apportent une représentation spatialisée de la productivité estimée à l'échelle des parcelles. Elles constituent un outil de visualisation essentiel pour relier les dynamiques spectrales et phénologiques observées aux performances agronomiques de la culture.

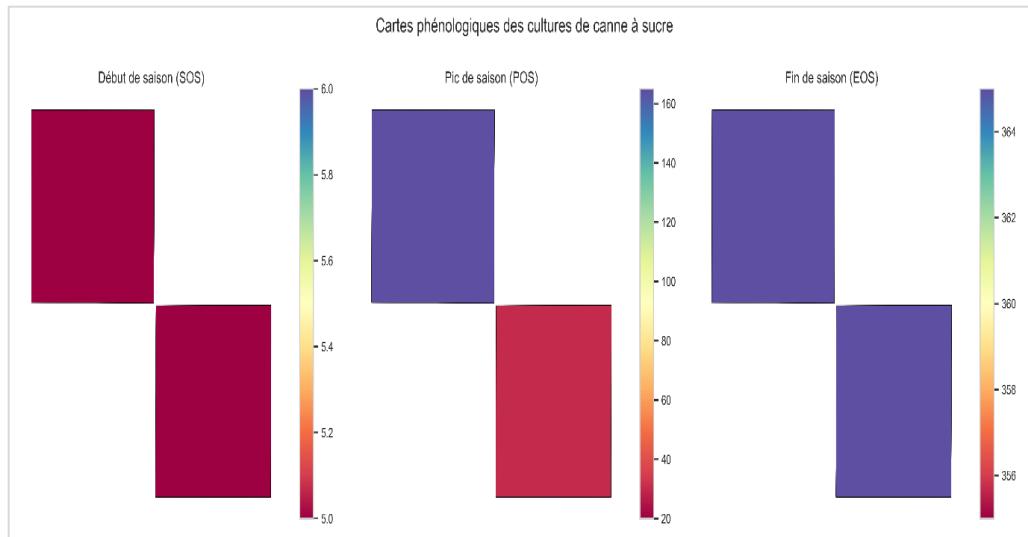


Figure 12. Cartes phénologique

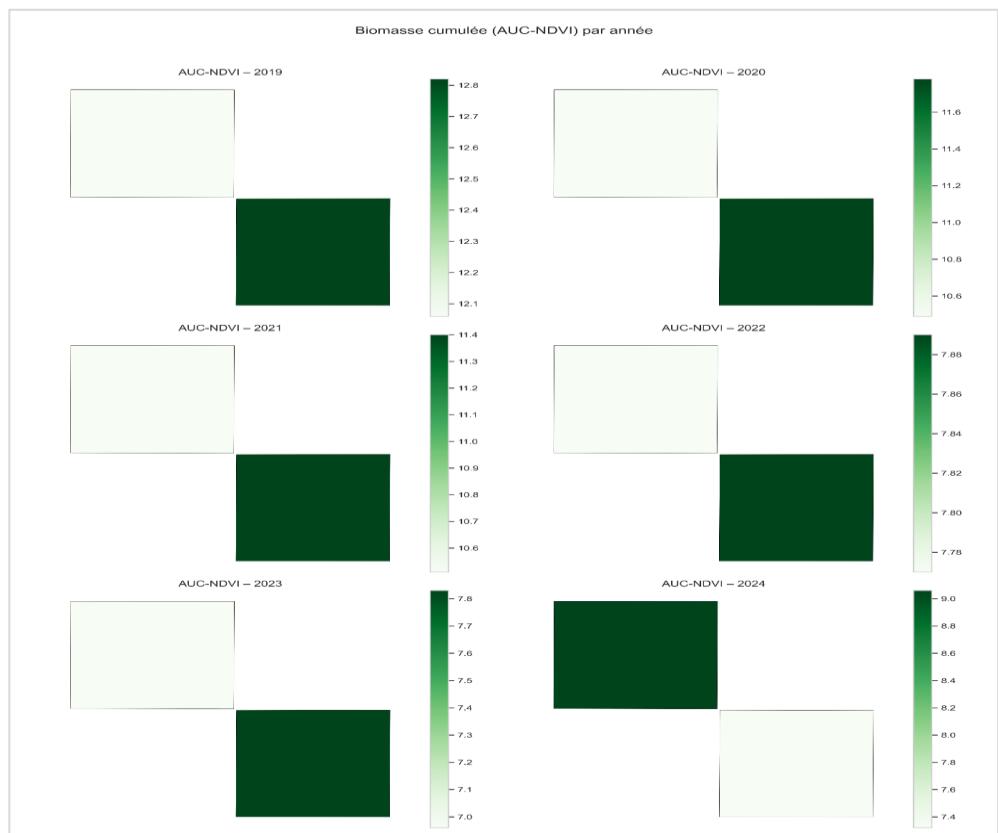


Figure 13. La carte AUC NDVI

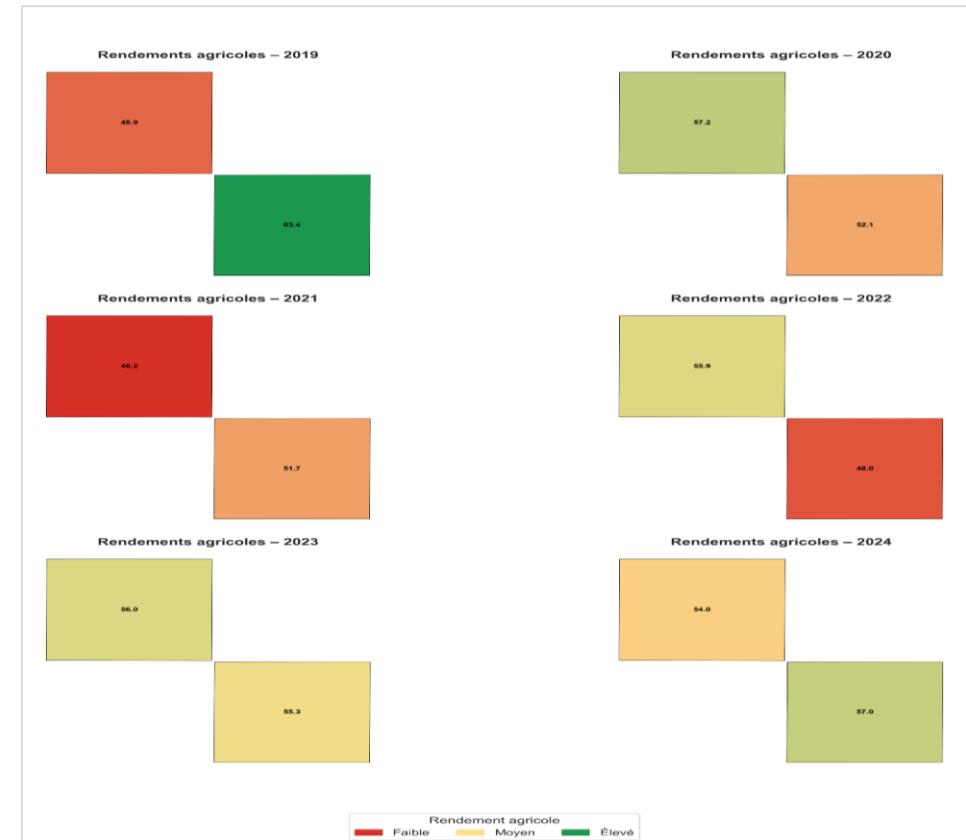


Figure 14. La carte des rendements prédit

3.2. DISCUSSION

L’analyse diachronique des images Sentinel-2 entre 2019 et 2024 a permis de caractériser finement la dynamique phénologique et les déterminants de la productivité de la canne à sucre en zone équatoriale. La série temporelle d’indices de végétation (NDVI, SAVI, EVI, MSAVI2, NDWI) a montré une forte cohérence avec le cycle cultural, révélant une saison de croissance longue (354–355 jours) et des phases distinctes de vigueur et de ralentissement. Ces résultats corroborent les observations de Bégué [7,9], qui ont démontré la pertinence des séries NDVI pour restituer la phénologie des cultures tropicales et anticiper leur rendement.

Les profils saisonniers obtenus mettent en évidence des pics de vigueur végétative en début d’année (janvier–mars), suivis d’un ralentissement entre août et octobre, période correspondant à une diminution de la photosynthèse active. Ce schéma, caractéristique des cycles de canne en zones tropicales humides, confirme que la croissance est fortement conditionnée par la disponibilité hydrique et la durée d’ensoleillement [29,30]. La corrélation positive entre NDVI et rayonnement solaire observée dans cette étude souligne le rôle central de la radiation dans l’accumulation de biomasse.

L’un des résultats majeurs réside dans la relation entre l’eau et la productivité. Le NDWI, corrélé positivement au rendement, confirme que la disponibilité hydrique est un facteur limitant essentiel. Des travaux antérieurs (Hemaprabha et al., 2004) avaient déjà montré qu’un déficit hydrique entraîne une réduction significative de la croissance et du rendement de la canne. Nos résultats renforcent cette conclusion et suggèrent que la résilience future de la culture dans un contexte équatorial dépendra largement de la gestion de l’eau, notamment face aux effets attendus du changement climatique sur la variabilité pluviométrique.

Sur le plan phénologique, l’importance du Start of Season (SOS) dans nos modèles prédictifs démontre que le démarrage précoce du cycle cultural est un facteur déterminant du rendement. Un SOS avancé permet d’optimiser l’interception lumineuse et de prolonger la phase de croissance active [31], en Afrique de l’Est, ont mis en évidence la même tendance, montrant que les variables phénologiques initiales expliquent une part importante de la variabilité interannuelle de la productivité. Cela confirme que l’optimisation des pratiques agricoles, telles que les dates de plantation, la préparation des sols et l’apport initial en fertilisants, reste un levier majeur pour sécuriser les rendements.

Sur le plan méthodologique, la comparaison entre régression multiple et Random Forest a révélé la supériorité du second ($R^2 = 0,818$). Ces résultats rejoignent ceux de Pandey et Taravat [29, 31], qui ont montré l’efficacité des approches d’apprentissage automatique pour modéliser les interactions complexes entre variables climatiques, spectrales et phénologiques. Toutefois, comme l’ont souligné Vianna et Nassif [32], ces modèles doivent être complétés par des approches mécanistes pour améliorer leur interprétation agronomique et guider les choix pratiques de gestion.

Enfin, la production de cartes phénologiques et de rendement prédit constitue une avancée opérationnelle majeure. Ces outils permettent d’identifier les zones à fort potentiel ainsi que les secteurs contraints, facilitant une gestion intra-parcellaire plus fine (irrigation ciblée, fertilisation différenciée). Ils ouvrent également la voie à la mise en place de systèmes d’alerte précoce, capables de signaler les baisses de productivité et de soutenir la résilience de la filière sucrière face aux aléas climatiques.

4. CONCLUSION

Cette recherche avait pour objectif d'explorer le potentiel des séries temporelles Sentinel-2, couplées à Google Earth Engine (GEE) et à des outils de modélisation en Python, pour analyser la phénologie de la canne à sucre et estimer ses rendements dans un contexte équatorial. Les résultats obtenus confirment la pertinence de cette approche : la saison culturelle a été caractérisée avec précision, les indicateurs phénologiques clés ont été extraits (SOS, POS, EOS, AUC-NDVI), et leur relation avec les variables climatiques a pu être établie. Ainsi, l'étude a atteint son objectif principal, qui était de proposer une méthodologie robuste et reproductible pour relier dynamiques spectrales, climat et productivité agricole.

Les résultats montrent que la canne à sucre présente un cycle long (354–355 jours) et fortement conditionné par la disponibilité en eau et le rayonnement solaire. La corrélation positive entre NDVI et ensoleillement, ainsi qu'entre NDWI et rendement, souligne l'importance des facteurs hydriques et radiatifs dans la croissance et l'accumulation de biomasse. De plus, l'influence majeure du Start of Season (SOS) dans les modèles prédictifs confirme que le déclenchement précoce du cycle culturel est un déterminant essentiel de la productivité. Ces observations rejoignent et renforcent les travaux antérieurs menés dans d'autres régions tropicales.

Au plan méthodologique, l'intégration d'outils de télédétection et d'apprentissage automatique constitue une avancée significative. La comparaison entre régression multiple et Random Forest a montré la supériorité de ce dernier ($R^2 = 0,818$), validant l'intérêt des approches non linéaires pour capturer la complexité des interactions entre climat, phénologie et rendement. Cette méthodologie, automatisée et transposable, illustre le potentiel des chaînes GEE–Python comme cadre analytique scalable, capable de traiter de grands volumes de données satellitaires et climatiques de manière efficace.

Sur le plan opérationnel, l'étude ouvre des perspectives concrètes pour l'agriculture de précision et la gestion des filières agricoles en contexte équatorial. La production de cartes phénologiques et de rendement prédit permet d'identifier les zones à fort potentiel et celles qui présentent des contraintes spécifiques, facilitant ainsi la mise en place d'interventions ciblées (irrigation, fertilisation, choix variétal). Ces outils offrent également une base pour la création de systèmes d'alerte précoce, capables de détecter les anomalies et d'anticiper les baisses de productivité. À terme, une telle approche pourrait contribuer à renforcer la résilience et la compétitivité de la filière sucrière congolaise, tout en servant de modèle pour d'autres cultures stratégiques dans la région.

Néanmoins, certaines limites doivent être reconnues. L'étude repose essentiellement sur des données optiques, sensibles à la couverture nuageuse, et n'intègre pas encore de manière exhaustive les informations relatives aux pratiques culturelles locales (intrants, dates de plantation, variétés cultivées) ni aux caractéristiques pédologiques. Ces lacunes restreignent la précision et la transférabilité des modèles. Il sera donc nécessaire, dans de futures recherches, d'intégrer des sources complémentaires, notamment l'imagerie radar (Sentinel-1), les données issues de drones ou CubeSats, ainsi que des observations de terrain plus détaillées.

En définitive, cette recherche démontre que la télédétection, couplée à l'intelligence artificielle, constitue un outil puissant pour comprendre, suivre et anticiper les dynamiques de la canne à sucre en zone équatoriale. Elle apporte une contribution à la fois scientifique, en améliorant la compréhension des liens entre phénologie, climat et rendement, et pratique, en fournissant des outils d'aide à la décision pour une gestion agricole durable. Dans un contexte de variabilité climatique croissante, la mise en œuvre de tels systèmes de suivi et d'alerte représente un levier majeur pour garantir la durabilité, la sécurité alimentaire et la compétitivité de la filière sucrière, et plus largement, pour soutenir une intensification agricole respectueuse des ressources naturelles.

REFERENCES

- [1] SARIS Congo. Rapport annuel d'activités. Nkayi: SARIS Congo; 2020.
- [2] FAO. L'agriculture en Afrique centrale : enjeux et perspectives. Rome: FAO; 2021.
- [3] Lobell DB, Asner GP. Cropland distributions from temporal unmixing of MODIS data. *Remote Sens Environ.* 2004;93(3):412–422. doi:10.1016/j.rse.2004.08.011.
- [4] Mutanga O, Skidmore AK. Narrow band vegetation indices overcome the saturation problem in biomass estimation. *Int J Remote Sens.* 2004;25(19):3999–4014. doi:10.1080/01431160310001654923.
- [5] Thenkabail PS, Lyon JG, Huete A, editors. *Hyperspectral Remote Sensing of Vegetation.* CRC Press; 2011.
- [6] Bégué A, Todoroff P, Paterne M. Modélisation du rendement de la canne à sucre à partir d'images satellitaires SPOT : application à La Réunion. CIRAD; 2010.
- [7] Bégué A, Lebourgeois V, Pellissier S. Suivi phénologique et cartographie de la récolte de la canne à sucre par télédétection à haute résolution spatiale. Rapport CIRAD, Projet SUCRETTE; 2015.
- [8] Duveiller G, Lopez-Lozano R, Seguini L. Crop monitoring using high temporal resolution remote sensing data: A case study in Europe. *Agric Syst.* 2015;135:1–12. doi:10.1016/j.agsy.2014.12.005.
- [9] Glover D, Jessup T. Strengthening agricultural decision-making with spatial data: Lessons from developing countries. *Dev Policy Rev.* 2019;37(2):231–249.
- [10] Gorelick N, Hancher M, Dixon M, Ilyushchenko S, Thau D, Moore R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens Environ.* 2017;202:18–27. doi:10.1016/j.rse.2017.06.031.
- [11] Orusa T, Viani A, Cammareri D, Borgogno Mondino E. A Google Earth Engine algorithm to map phenological metrics in mountain areas worldwide with Landsat Collection and Sentinel-2. *Geomatics.* 2023;3(1):221–238. doi:10.3390/geomatics3010012.
- [12] Roy DP, Wulder MA, Loveland TR, et al. Landsat-8: Science and product vision for terrestrial global change research. *Remote Sens Environ.* 2014;145:154–172. doi:10.1016/j.rse.2014.02.001.
- [13] Main-Knorn M, Pflug B, Louis J, et al. Sen2Cor for Sentinel-2 — ESA Level-2A processor. *Proc SPIE Image Signal Process Remote Sens.* 2017;10427:1042704. doi:10.1117/12.2278218.
- [14] Google Earth Engine Community. Sentinel-2 Cloud Masking with s2cloudless (tutorial). Google Developers; last updated 2025-09-29. <https://developers.google.com/earth-engine/tutorials/community/sentinel-2-s2cloudless>
- [15] Baetens L, et al. Validation of Copernicus Sentinel-2 cloud masks obtained from MAJA, Sen2Cor and FMask using reference cloud masks generated with a supervised active-learning procedure. *Remote Sens.* 2019;11(4):433. doi:10.3390/rs11040433.
- [16] De Simone L, Gennari P. Earth observations for official crop statistics in the context of scarcity of in-situ data. *Stat J IAOS.* 2022;38(3):1009–1019. doi:10.3233/SJI-220054.
- [17] Rouse JW, Haas RH, Schell JA, Deering DW. Monitoring vegetation systems in the Great Plains with ERTS. In: Third ERTS Symposium Proceedings. NASA SP-351; 1974:309–317.
- [18] Huete AR, Justice CO, et al. Overview of MODIS vegetation indices (EVI). *Remote Sens Environ.* 2002;83(1-2):195–213. doi:10.1016/S0034-4257(02)00096-2.
- [19] Qi J, Chehbouni A, Huete AR, Kerr YH, Sorooshian S. A modified soil adjusted vegetation index (MSAVI2). *Remote Sens Environ.* 1994;48(2):119–126. doi:10.1016/0034-4257(94)90134-1
- [20] Gao BC. NDWI A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sens Environ.* 1996;58(3):257–266. doi:10.1016/S0034-4257(96)00067-
- [21] Misra G, et al. Status of phenological research using Sentinel-2 data: a review. *Remote Sens.* 2020;12(17):2760. doi:10.3390/rs12172760.
- [22] Liang J, et al. Using enhanced gap-filling and Whittaker smoothing to reconstruct high spatio-temporal NDVI time series based on Landsat 8, Sentinel-2 and MODIS imagery. *ISPRS Int J Geo-Inf.* 2023;12(6):214. doi:10.3390/ijgi12060214.

- [23] White MA, de Beurs KM, et al. Intercomparison, interpretation and assessment of spring phenology estimated from remote sensing for 1982–2006. *Global Change Biol.* 2009;15(10):2335–2359. doi:10.1111/j.1365-2486.2009.01910.x.
- [24] Barnett V, Lewis T. *Outliers in Statistical Data*. John Wiley & Sons; 1994.
- [25] Benesty J, Chen J, Huang Y, Cohen I. Pearson correlation coefficient. In: *Noise Reduction in Speech Processing*. Springer; 2009. p.1–4. doi:10.1007/978-3-642-00296-0_5.
- [26] Breiman L. Random forests. *Machine Learning*. 2001;45(1):5–32. doi:10.1023/A:1010933404324.
- [27] Willmott CJ, Matsuura K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Clim Res.* 2005;30:79–82. doi:10.3354/cr030079.
- [28] Defourny P, Bontemps S, et al. Near real-time agriculture monitoring at national scale at parcel resolution: Performance assessment of the Sen2-Agri automated system. *Remote Sens Environ.* 2019;doi:10.1016/j.rse.2018.11.007.
- [29] Pandey S, Patel NR, Danodia A, Singh R. Discrimination of sugarcane crop and cane yield estimation using multi-temporal satellite data. *ISPRS Archives*; 2019.
- [30] Hemaprabha G, Nagarajan R, Alarmelu S. Response of sugarcane genotypes to water deficit stress. *Sugar Tech.* 2004;6(3):165–168. doi:10.1007/BF02942718.
- [31] Taravat A, Abebe G, Gessesse B, Tadesse T. Estimation of sugarcane yield using multi-temporal Sentinel-2 satellite imagery and Random Forest regression. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* 2024;48:357–362.
- [32] Vianna MDS, Nassif DSP, Carvalho KS, Marin FR. Modelling the trash blanket effect on sugarcane growth and water use. *Comput Electron Agric.* 2020;172:105361. doi:10.1016/j.compag.2020.105361.