Design Project

Mesurer la biomasse ligneuse en zone tropicale

Étudiants

Baretje Aurèle Kamoun Ines

Superviseur EPFL

Devis Tuia

Collaborateur ArboRise

Philippe Nicod

Lausanne, 7 Juin 2024

Contexte & question scientifique

ArboRise est une ONG basée à Lausanne, dédiée à la reforestation en Guinée, plus précisément en Haute Guinée et en Guinée forestière). Son objectif principal est de restaurer le couvert forestier de la région grâce à une collaboration avec la population locale. À ce jour, arboRise doit surveiller la croissance de plus de 600 parcelles, chacune ayant en moyenne une taille de deux hectares. La surveillance et l'estimation de la croissance de la végétation ligneuse sont essentielles pour comprendre le succès ou l'échec des projets de reforestation. En effet, mener une surveillance sur le terrain autour de 26 villages avec des parcelles de tailles variées présente des défis significatifs en terme de coûts et de temps. Par ailleurs, plus de 600 parcelles sont distribuées sur une région d'environ 960km^2.
L'objectif de ce projet est d'utiliser des techniques de télédétection pour quantifier la croissance de la biomasse ligneuse et, par conséquent, la croissance des arbres. Cette étude initiale vise à distinguer les parcelles dont la reforestation a réussie (ou est en cours de réussite) des parcelles ou cela n'a pas fonctionné. Puis le second objectif est d'identifier les facteurs naturels et anthropiques communs parmi les parcelles où la reforestation a réussie/échoué. Les caractéristiques naturelles telles que les propriétés du site ou les types de graines, ainsi que les facteurs anthropiques tels que la proximité des routes et des villages, peuvent jouer des rôles cruciaux. La proximité des routes et des villages sont par exemple liées à la présence de pâturage et le potentiel passage par les parcelles des animaux, [Koita, 1998].

1 Introduction

Diverses méthodologies sont communément employées pour estimer la biomasse aérienne (AGB) des arbres en utilisant des techniques de télédétection [Shi and Liu, 2017]. Une première approche repose sur l'estimation de la hauteur des arbres à l'aide de technologies radar ou de détection et de télémétrie par ondes lumineuses (LiDAR) [Potapov et al., 2021]. De plus, certaines études ont démontré des corrélations entre l'AGB et des indices de végétation traditionnels tels que l'indice de végétation par différence normalisée (NDVI) [Askar et al., 2018].

La première méthode, telle qu'elle est décrite, repose sur l'estimation de la hauteur des arbres en intégrant des technologies radar ou LiDAR. Par exemple, l'instrument GEDI (Global Ecosystem Dynamics Investigation) équipé de trois lasers permet d'estimer divers paramètres forestiers, dont la hauteur de la canopée et la biomasse depuis l'espace. L'association de ces données avec celles du radar à synthèse d'ouverture (SAR) (comme l'imagerie Sentinel-1) et d'autres satellites multispectraux (comme Sentinel-2) offre un potentiel important pour l'estimation de la hauteur de la canopée sur de vastes zones géographiques [Lang et al., 2022]. Ensuite à partir de cette hauteur, une estimation de l'AGB peut-être réalisée.
Cependant, les modèles générés à partir des données de GEDI ne sont pas précis pour la végétation d'une hauteur inférieure à 3 ou 5 mètres puisque la marge d'erreur des modèles est d'environ 3 mètres. Par conséquent, utiliser cette approche dans des contextes impliquant des jeunes arbres récemment plantés, dont les hauteurs prévues n'excèdent pas 5 mètres comme pour la végétation des parcelles arboRise, n'est pas réaliste puisque l'incertitude ne permet pas de distinguer une croissance de quelques dizaines de centimètres. De plus, le nombre de passage de GEDI au dessus de la zone d'intérêt est faible et n'aurait pas permis de couvrir toute la zone d'étude sur la période qui nous intéresse. Dans de tels cas, d'autres méthodologies mieux adaptées aux stades de croissance et aux caractéristiques spécifiques de la végétation doivent être envisagées pour une estimation précise de la biomasse.

Une autre méthode d'estimation de la biomasse consiste à utiliser des indices de végétation tels que le NDVI. Ces indices spectraux fournissent des indications sur la productivité et la photosynthèse, et peuvent être utilisés pour estimer la biomasse. Cependant, l'utilisation du NDVI est généralement moins précise et plus variable, dépendant de l'emplacement et du moment des mesures. Des facteurs tels que l'inclinaison, la couverture nuageuse et le moment des mesures doivent être pris en compte pour comparer les valeurs de NDVI, et il peut-être difficile de distinguer la biomasse ligneuse (les arbres) de la biomasse totale, comprenant la végétation non ligneuse (plantes herbacées, buissons, etc).
Pour dériver la biomasse aérienne à partir du NDVI, il est nécessaire de développer un modèle utilisant des données de terrain sur la biomasse ou la productivité réelle de parcelles spécifiques [Foody et al., 2001]. Dans notre cas, les données dendrométriques de terrain n'étaient pas suffisantes pour obtenir une correspondance entre le NDVI et la biomasse réelle sur le terrain. De plus, la région étudiée se trouve dans une zone tropicale avec une saison des pluies importante et une couverture nuageuse élevée pendant environ 6 mois rendant les mesures de NDVI impossibles.

Il nous est donc impossible de quantifier l'AGB (Above-Ground Biomass, biomasse aérienne) pour évaluer la croissance des parcelles arboRise. Puisqu'estimer l'augmentation de la hauteur des plantes ou évaluer la biomasse aérienne à partir des valeurs NDVI ne convient pas dans le cas des nouvelles plantations à faible hauteur et lorsque les données de terrain sont limitées, nous avons par conséquent opté pour un système de clustering afin d'estimer si une parcelle est en croissance ou non. L'objectif n'étant plus de quantifier la biomasse ligneuse mais d'évaluer si cette biomasse est en croissance ou pas. Ensuite sur la base de cette classification, nous pourrons nous focaliser sur les facteurs anthropiques et naturels qui pourraient favoriser la croissance ou la non croissance de certaines parcelles.

2 Sources de données

2.1 Données nécessaires à l'estimation de la croissance

Nous disposons pour la réalisation de ce projet des données relatives aux parcelles arboRise, comprenant des données géospatiales (morphologie, localisation et surface des parcelles), ainsi que des méta-données comprenant l'année de reforestation, les propriétaires des parcelles, les villages correspondants, etc. De plus, nous disposons d'une description qualitative de l'évolution de 20 parcelles effectuée en 2023. Cette description n'est plus à jour pour l'année 2024 puisque des évènements tels que des feux ont eu lieu et ont ralenti la croissance de certaines parcelles.
Dans toute cette étude nous nous sommes focalisés sur les séries temporelles du NDVI afin d'évaluer la croissance de la biomasse des parcelles. Cet indice NDVI permet de mettre en évidence la quantité de la végétation. Nous utilisons des données fournies par le satellite Sentinel-2 pour les années 2021, 2022, 2023 et 2024, plus précisément, le NDVI est calculé à partir des bandes "Red", band 3, et "NIR", band 8, de Sentinel-2. Nous avons spécifiquement privilégié les bandes spectrales capturées par Sentinel pour le mois de février, une période de saison sèche en Guinée, correspondant à une couverture nuageuse faible (nous utilisons uniquement les jours avec une couverture nuageuse $<5\%$) ce qui permet d'avoir plus de données avec Sentinel (Table 6).
Le choix de prendre les données de la saison sèche permet aussi de s'affranchir le plus possible des effets de la végétation non ligneuse (herbacée principalement) puisque celle-ci contribue peu au NDVI en saison sèche [Lebrun, 2022] et qu'elle ne nous intéresse pas pour notre étude.
Le choix d'utiliser uniquement le mois de février a été fait à cause de la variabilité du NDVI dans le temps, qui introduit du bruit dans les données. L'ajout des mois de janvier et/ou de mars ne permet pas de préciser nos analyses, au contraire il masque l'évolution au fil des ans dans la variabilité inter-mensuelle observée entre les mois de janvier et mars. Cette variabilité semble inhérente à notre méthodologie et à l'utilisation du NDVI et s'observe aussi dans la littérature [Vo Quang Truong, 2022].

2.2 Sources de données relatives aux facteurs anthropiques et naturels

Les facteurs anthropiques et naturels utilisés pour évaluer un potentiel impact sur la croissance des parcelles ont été divisés en trois catégories : les facteurs abiotiques, les perturbations naturelles ou anthropiques, et le mode de gestion et de plantation (facteurs biotiques). Une analyse détaillée de ces facteurs et des motivations de nos choix sera présentée dans la partie relative aux méthodologies. Dans cette section, nous allons présenter les sources des données utilisées.

2.2.1 Facteurs abiotiques et géographiques

Caractéristiques du sol : Nous avons récupéré les données relatives aux propriétés du sol pour les parcelles reboisé via une cartographie numérique mondiale des sols avec une résolution de 250 mètres [SoilGrids, 2024]. Nous avons extrait les propriétés physiques du sol, telles que la concentration massique d'argile (en g/kg), de sable (en g/kg), et de limon (en g/kg), pour la couche de sol allant de 5 à 15 cm de profondeur. De plus, nous avons extrait des propriétés chimiques telles que le pH du sol, la concentration massique d'azote (en cg/kg), la teneur du sol en matière organique (dg/kg) et la capacité d'échange cationique à pH 7 (mmol(c)/kg). La table 1 représente certaines notations qui seront utilisées pour certaines figures dans la partie relative aux résultats.

L'altitude, la pente et l'exposition ont été calculées à partir d'un DEM "Digital Elevation Model" qui fournit l'altitude dans la région (*Altitude*). Nous avons utilisé le DEM GLO-30 de Copernicus [Copernicus, 2021]

Facteur	*clay 5-15*	*cec 5-15*	*cfvoc 5-15*	*soc 5-15*	*sand 5-15*	*silt 5-15*	*nitrogen 5-15*	*soil pH 5-15*
Définition	teneur en argile ($<$ $0.002mm$ de diamètre)	capacité d'échange cationique à pH 7	fraction volumique des fragments ($>$ $2mm$ de diamètre)	teneur du sol en matière organique	teneur en sable ($>$ $0.05mm$ de diamètre)	teneur en limon ($>$ $0.002mm$ et $<$ $0.05mm$ de diamètre)	teneur en azote	pH du sol multiplié par 10

Table 1: Notations des facteurs de sol

ayant une résolution spatiale de 30m. Les calculs pour obtenir la pente moyenne des terrains, et l'écart type, (*mean_pente_terrain* et *std_pente_terrain*) ainsi que l'exposition (*mean_aspect* et *std_aspect*) ont été faits à l'aide du logiciel QGIS. On a aussi transformé le facteur *mean_aspect* en un facteur qualitatif *Exposition* comprenant la direction de l'exposition par rapport aux points cardinaux (N, NW, W, SW, S, SE, E, NE).

La **latitude** (*lat*) et la **longitude** (*lon*) ont été obtenues en prenant le centroïde des parcelles.

2.2.2 Perturbations naturelles ou anthropiques

Localisation des feux : Nous avons extrait la localisation des feux autour des parcelles grâce au système FIRMS (Fire Information for Resource Management System) fourni par la NASA pour les années allant de 2021 à 2024 [NASA, 2024]. Avec ces données nous avons extrait la distance au feu le plus proche, pour chaque années, en utilisant la fonction "Join attributes by nearest". Nous avons ensuite pris la distance minimale des feux extraits entre l'année suivant la plantation et 2024 (*min_dist_feu*). De même nous avons pu extraire le nombre de feux dans un rayon de 500m et 100m autour des parcelles (*feux_2024_500* et *feux_2024_100*).

Distance à la route, distance au village Les routes ont été définies manuellement à partir des données satellites. Nous avons choisi d'établir le tracé des routes de cette manière afin d'inclure les chemins empruntés par les motos et le bétail, et pas uniquement les grandes voies officielles. Ces petits chemins sont couramment utilisés par la population locale pour se déplacer d'une région à une autre. Pour les villages, nous avons récupéré des polygones représentant la localisation des villages dans la région d'intérêt. Nous avons utilisé le logiciel de système d'information géographique QGIS pour déterminer la distance entre chaque parcelle et la route ou le village le plus proche (*distance_village*, *distance_parcelle* et *distance_routes_chemins*).

2.2.3 Mode de plantation (facteurs biotiques) et de gestion

Les informations relatives à la localisation des parcelles, au village gestionnaire (*Village*), à l'année de plantation (*Année*) et au groupe de plantation (*Phase*) font partie des méta-données fournies par arboRise. Le groupe de plantation est divisé en 4 catégories. Une catégorie "pilote" qui correspond au première parcelles reforestées en 2021. Les 3 autres catégories sont "le groupe 1", "le groupe 2" et "le groupe 3". Chaque groupe représente un ensemble de villages pour lesquels des espèces avec une période de maturité différente sont plantées (groupe 1 : janvier-février, groupe 2 : mars-avril et groupe 3 : mai-juin).
Les données de distance à la forêt la plus proche (*distance_foret*) ont été obtenues à l'aide de QGIS et du jeu de données ALOS [ALOS, 2020]. Les données relatives à la distance à l'eau la plus proche (*distance_eau*) ont été obtenues avec QGIS et la base de données de la FAO [AQUASTAT, 2021].
Les données relatives à la baseline ont été obtenues à partir du NDVI des parcelles au 14/02/2021 comme décrit ci-après (Méthodologie 3.3.1).

3 Méthodologie

3.1 Pré-traitement des données

Les parcelles reboisées par arboRise présentent une certaine variabilité concernant la végétation présente avant la reforestation. Certaines parcelles sont totalement nues, tandis que d'autres contiennent quelques arbustes ou même des arbres matures. Afin d'évaluer la croissance attribuable au projet arboRise, nous avons supprimé les pixels représentant des arbres matures. Ces arbres se caractérisent par une valeur de NDVI importante, nettement supérieure à celle des arbustes ou de l'herbe environnante.
Nous nous sommes focalisés sur les valeurs du NDVI du 14/02/2021 afin de distinguer les arbres matures du reste de la végétation dans les parcelles du projet. Le choix de cette date correspond à une période antérieure à la première reforestation réalisée en 2021. Ainsi, nous avons utilisé le NDVI du 14/02/2021 pour supprimer les arbres de toutes les parcelles reboisées en 2021, 2022 et 2023 afin de s'assurer de l'uniformité du critère de suppression.
En établissant un seuil du NDVI de 0,5 pour caractériser les arbres, nous avons supprimé les parties de parcelles correspondantes à cette végétation comme le montre la figure 1 (nous avons supprimé des parcelles les pixels oranges). Le choix de ce seuil a été déterminé en testant plusieurs valeurs et en évaluant si la localisation des pixels supprimés correspondait à des arbres, sur la base des images satellites.
Dans la suite de notre étude, nous allons évaluer le NDVI des parcelles après suppression des arbres matures. Grâce à ce pré-traitement nous pouvons ainsi travailler sur des parcelles exemptes d'arbres matures pour évaluer la croissance des plantations de arboRise. Nous réintroduirons par la suite la baseline comme facteur potentiel d'influence de la croissance (on peut penser aux effets des graines, de l'ombrage, des parasites, de l'humidité, des mycorhizes, de la quantité de nutriments, etc)

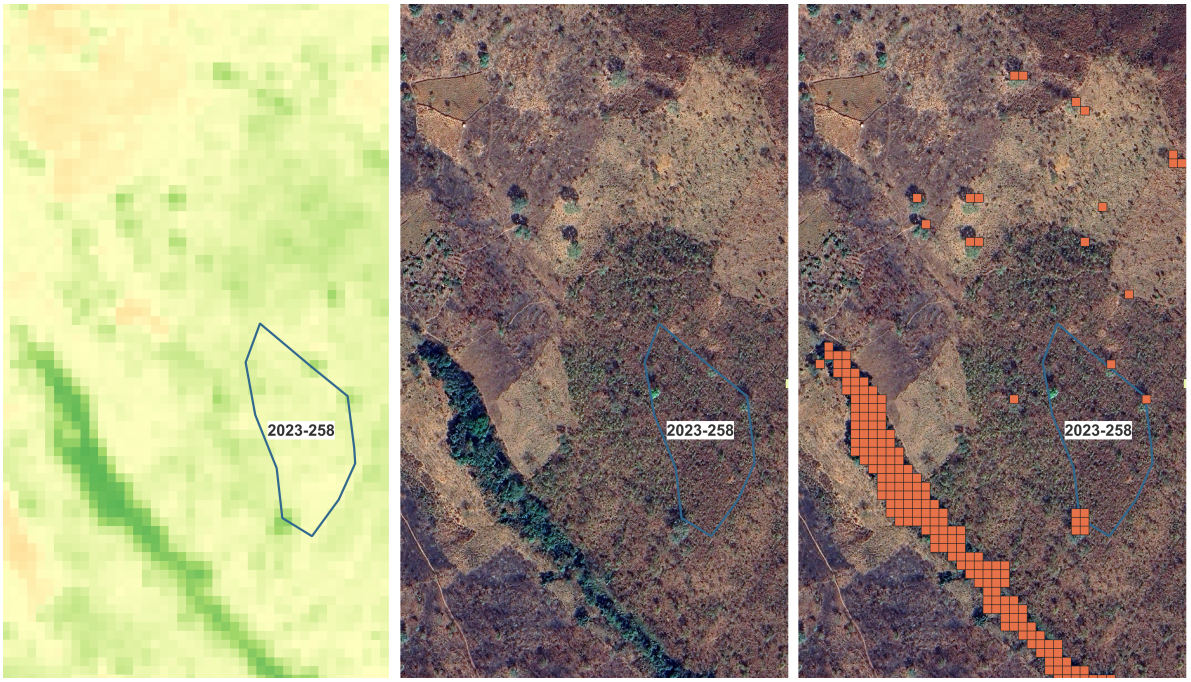

Figure 1: Image satellite du NDVI obtenue à partir de Sentinel-2 au 14/02/2021 sur la parcelle "2023-258" et ses environs. Image satellite provenant de Google Satellite (en date du 21/01/2024) sur la même zone. Image satellite avec, en orange, les pixels ayant une valeur de NDVI supérieure à 0,5.

3.2 Clustering des parcelles

Puisque, par manque de données de terrain, nous ne pouvons pas quantifier exactement l'AGB de chaque parcelle, nous avons opté pour l'utilisation du clustering, une méthode non supervisée, pour catégoriser les parcelles. Par ailleurs, nous avons adopté une approche comparative en employant deux modèles distincts et en évaluant la

cohérence de leurs clusterings comme méthode de validation.
Dans cette partie, nous allons présenter les deux méthodes de clustering distinctes utilisées pour cette étude. La première méthode, que nous appellerons "méthode 1", évalue la performance absolue de chaque parcelle dans le temps. La deuxième méthode, que nous appellerons "méthode 2", évalue la performance de chaque parcelle relativement à la performance des alentours.

3.2.1 Methode 1 : clustering en utilisant les données à l'échelle des pixels

Dans notre première méthode, chaque parcelle a été subdivisée en une grille de pixels de dimensions 64 pixels x 64 pixels. À chaque pixel, nous associons une série temporelle des valeurs du NDVI sur la période allant de l'année de reforestation de la parcelle jusqu'en 2024. Par conséquent, le nombre de valeurs de NDVI est réduit pour les parcelles reboisées en 2023 et la série temporelle est plus courte (voir Table 6 pour le détail des dates utilisées).
En utilisant une approche de régression linéaire, nous calculons la pente de la régression linéaire de la série temporelle pour chaque pixel, reflétant ainsi la tendance des valeurs de NDVI [Lambert et al., 2015]. Nous obtenons ainsi un nouveau raster représentant les pentes des régressions linéaires pour chaque parcelle et donc la variation temporelle du NDVI pour chaque pixel.
En utilisant une régression linéaire, nous atténuons la variabilité du NDVI entre chaque jour du mois, nous permettant ainsi de nous concentrer sur la tendance de variabilité du NDVI. La figure 2 représente les valeurs des pentes de la régression linéaire pour la parcelle 2022-063 ainsi qu'un exemple de régression pour un pixel en particulier.

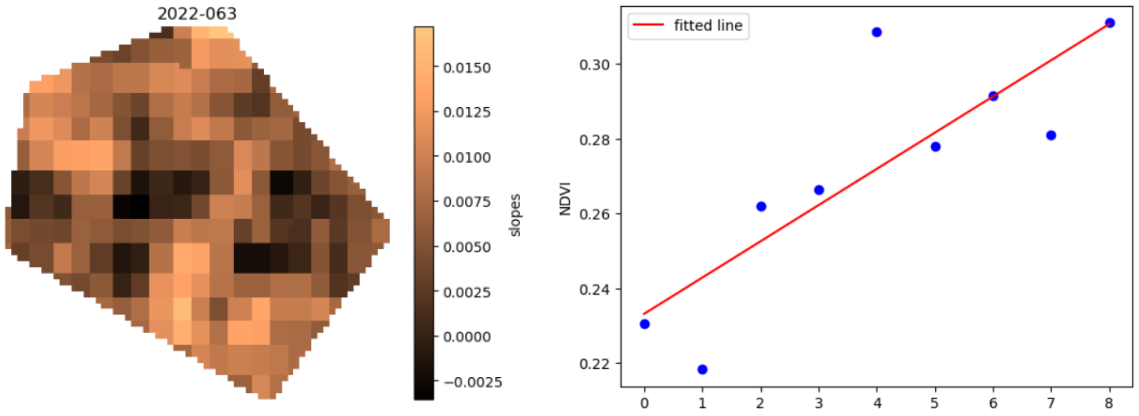

Figure 2: Valeur des pentes de la régression linéaire pour la parcelle 2022-063. La pente pour chaque pixel est récupérée en utilisant une régression linéaire comme représentée sur la figure de droite. Les pixels clairs correspondent à des parties de la parcelle dont la biomasse augmente le plus, et les pixels sombres à des parties de la parcelle dont la biomasse augmente pas ou peu (voir diminue).

Pour classifier les parcelles, nous utilisons l'algorithme de clustering k-means. La caractéristique des parcelles intégrée dans l'algorithme de k-means pour effectuer le clustering est la moyenne des pentes sur tous les pixels de chaque parcelle. Nous avons décidé de ne pas prendre en compte d'autres facteurs dans le clustering, comme la variabilité de la croissance entre les pixels, afin de conserver un critère commun entre cette méthode et la méthode 2 présentée ultérieurement.
L'algorithme k-means permet de trouver les clusters minimisant la variabilité intra-clusters tout en maximisant la variabilité inter-clusters, ce qui permet de trouver les clusters "naturels" plus simplement. Nous avons d'abord divisé les données en deux clusters, puis une deuxième fois en trois clusters.
Outre le fait qu'elle permet une classification de la performance de chaque parcelle, cette méthode permet également à arboRise de transmettre des consignes de correction de performance aux propriétaires des parcelles.

3.2.2 Méthode 2 : clustering en utilisant les données moyennées et normalisées

Dans notre seconde méthode, nous utilisons une valeur relative du NDVI. Concrètement, nous calculons la différence entre le NDVI moyen de chaque parcelle et le NDVI moyen du centre du village le plus proche pour chaque date (Figure 3). Nous prenons le centre du village le plus proche comme référence, car la valeur du NDVI de ce dernier

est relativement stable au cours des années en lien avec l'occupation de cet espace caractérisé par une végétation très faible voir nulle. Cela permet de mieux saisir l'augmentation intrinsèque du NDVI et donc de l'activité photosynthétique et de la biomasse des parcelles. En effet l'évolution de l'indice ainsi mesuré est donc une évolution relative qui prend en compte une partie de la variabilité du NDVI. Il est important de choisir comme référence une zone dont le NDVI varie peu, sinon on risque de mesurer l'évolution de la biomasse du point de référence et pas celle de notre parcelle. C'est aussi une des raisons qui ont porté notre choix sur les centres des villages pour servir de valeur de référence dans cette méthode. Comme dans la première méthode nous effectuons une régression linéaire de la différence obtenue afin de prendre en compte l'évolution à long terme.
De la même manière la série temporelle considérée se base uniquement sur les données du mois de février de 2021 à 2024 (Table 6). Pour les parcelles plantées en 2022 (2023), nous ne prenons que les valeurs de NDVI à partir de 2022 (respectivement 2023). Les calculs sur les parcelles de 2023 reposent donc sur un nombre faible de données (5 dates) par rapport aux parcelles de 2022 et 2021.
Comme dans la méthode précédente, un k-means est effectué pour classifier les parcelles, en deux puis trois clusters.

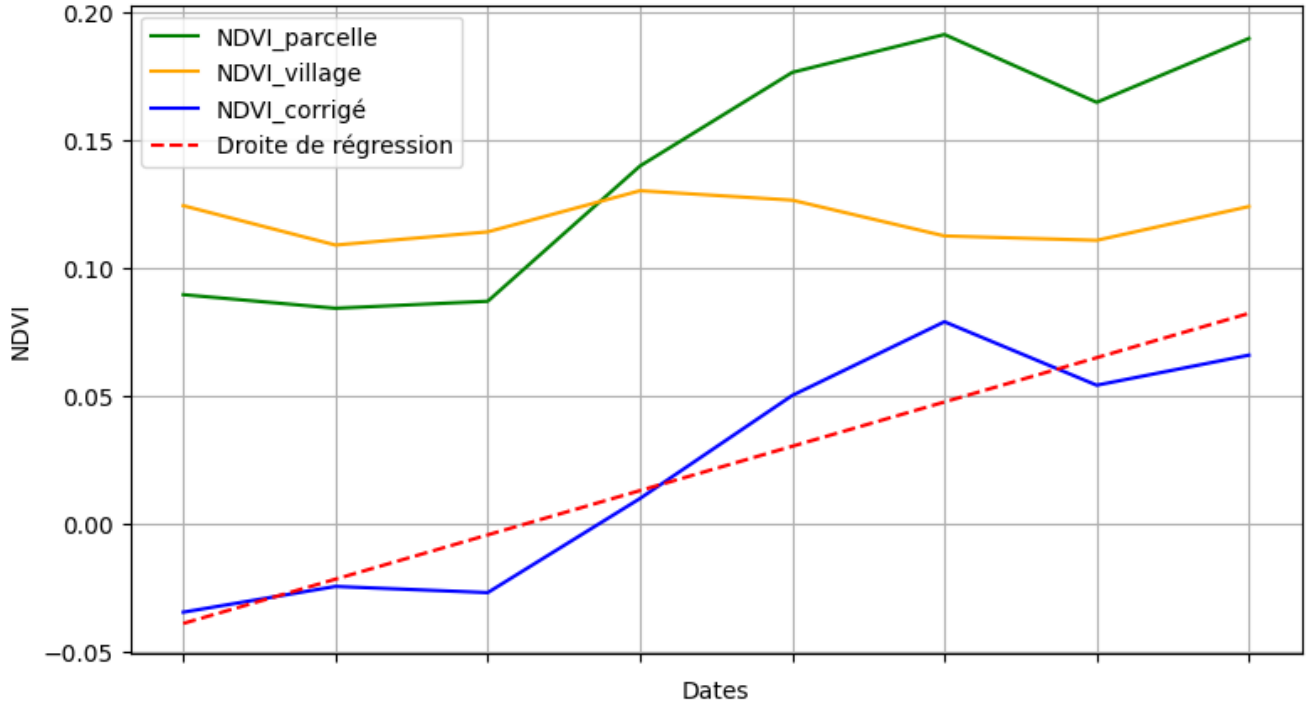

Figure 3: Valeur de NDVI (entre 2022 et 2024) pour la parcelle 2022-186 (NDVI_parcelle), pour le village (NDVI_village) correspondant (Tiegbenina) et NDVI corrigé (NDVI_corrigé), correspond à la différence du NDVI de la parcelle et de celui du village. Pour cette valeur corrigée du NDVI, la droite de régression est également ajoutée ici en rouge.

3.2.3 Notation du clustering des parcelles

Nous avons entrepris un clustering avec deux et trois clusters distincts. Pour le cas à deux clusters, nous avons assigné les parcelles à deux catégories : "bien" pour le cluster ayant une pente de régression la plus élevé, donc les parcelles qui manifestent une "bonne" croissance ; et "pas bien" pour le cluster avec des valeurs faibles de la pente de régression (parfois négative), donc les parcelles qui ne présentent pas ou peu de signes de croissance. En ce qui concerne la clustering à trois clusters, nous avons défini les clusters comme "bien", "moyen" et "mauvais" de la même manière. L'analyse de ces clustering à 2 ou 3 catégories sera présentée dans la section 4.1.

3.3 Facteurs anthropiques et naturels

Pour choisir et identifier les facteurs pouvant influencer la croissance de la biomasse ligneuse, nous avons procédé à une analyse systémique réalisée en collaboration avec arboRise pour tenir compte des réalités sur le terrain. On peut postuler que ces facteurs se regrouperont en trois catégories :

- **Les facteurs abiotiques (physico-chimiques) et géographiques** incluent les caractéristiques du sol qui influencent la croissance des plantes [Dang et al., 2022], ainsi que les facteurs de localisation, d'altitude et d'exposition [Kasekete, 2024].
- **Les perturbations naturelles ou anthropiques** comprennent les feux, ainsi que la localisation des parcelles par rapport aux villages et aux routes. Une proximité avec ces éléments peut faciliter l'accès à ces terrains par le pâturage ou, au contraire, favoriser la protection de la zone par les gestionnaires. [Ouattara et al., 2022]
- **Les modes de plantation (facteurs biotiques) et de gestion** incluent les informations relatives au village gestionnaire, l'année de plantation et la baseline de la parcelle et des alentours pouvant influencer plus directement la croissance.

D'autres facteurs auraient pu être considérés mais nous avons fait le choix (et été en partie contraints) de renoncer à certains d'entre eux (Discussion 5.4).

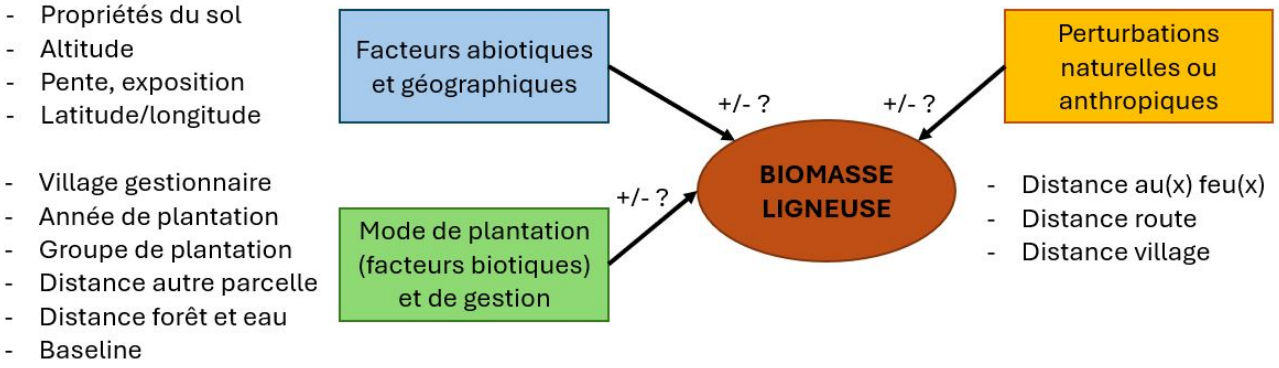

Figure 4: Facteurs pouvant influencer la croissance de la biomasse ligneuse.

3.3.1 Méthode d'estimation de la baseline

La suppression des arbres matures dans les parcelles lors de la phase de pré-traitement nous a permis de prendre en compte la variation du NDVI due à la reforestation effectuée par arboRise, en éliminant la contribution des arbres matures. En revanche, la végétation buissonnante et les arbustes, par exemple, n'ont pas été supprimés ni pris en compte dans notre étude.
Cette baseline (végétation buissonnante et arbustes) peut être considérée comme un facteur influençant la croissance des parcelles. En effet, elle peut créer de l'ombre, influencer l'humidité de la parcelle, favoriser la croissance via les relations mycorhiziennes, ou la réduire en attirant des parasites et des prédateurs.
Pour prendre en compte et évaluer cette baseline (végétation buissonnante et arbuste, facteur : *Baseline*) sur les parcelles, nous avons utilisé les données de NDVI ainsi qu'une méthode de clustering pour obtenir une estimation approximative de la quantité de végétation présente avant la reforestation sur chaque parcelle. Afin de permettre de bonnes comparaisons entre les parcelles, nous avons choisi de nous baser sur les valeurs du NDVI de 2021, plus particulièrement sur les données du 14 février 2021 comme cela a été fait plus tôt. Nous avons extrait pour chacune des parcelles l'histogramme des valeurs NDVI des pixels ne contenant pas d'arbres mature (NDVI < 0.5). À partir de cet histogramme, un clustering en trois clusters a été réalisé avec la méthode k-means. Le cluster avec le NDVI moyen le plus élevé est appelé baseline "forte", le cluster avec le NDVI moyen le plus faible est appelé baseline "faible", le cluster restant correspond à la baseline "moyenne".

De plus nous avons créé un nouveau facteur quantitatif (*prop_big_trees*) qui correspond à la proportion que représente les arbres matures par rapport à la surface totale de la parcelle. Il s'agit donc d'un nombre entre 0 et 1 qui quantifie la baseline en arbres matures de chaque terrain.
Nous aboutissons donc à deux facteurs relatifs à la baseline ayant une influence potentielle sur la croissance : le premier est lié la végétation buissonnante et les arbustes, et le deuxième est rattaché aux arbres matures sur la parcelle.

3.3.2 Contribution des facteurs quantitatifs

Pour identifier les facteurs jouant les rôles les plus importants dans la définition des clusters ("bien", "pas bien"), nous avons utilisé l'outil : "Caractérisation d'une variable qualitative" du package Factoshiny sur RStudio. Cet outil permet comme son nom l'indique de caractériser une variable qualitative (ici nos clusters de parcelles, "bien"/"pas bien" ou "bien/mauvais"). Il permet de trouver quels sont les facteurs qui sont les plus corrélés à notre classification. Par exemple un facteur quantitatif pour lequel beaucoup de parcelles "bien" sont associées à des fortes valeurs, et beaucoup de parcelles "pas bien" sont associées à des faibles valeurs pourra ressortir de cette analyse et aura une influence positive sur notre classification. Des facteurs n'ayant pas d'influence, autant de fortes ou faibles valeurs dans les 2 groupes, ne ressortiront pas de cette analyse.
Nous avons effectué l'analyse à plusieurs reprises avec différents sous-jeux de données (Tables 4 & 5), en gardant à chaque fois uniquement les parcelles dont la classification concorde entre les deux méthodes précédemment décrites. Cela permet de s'assure de la robustesse des facteurs (ce n'est pas juste à cause de l'effet d'une année ou de quelques parcelles "moyennes" que les facteurs ressortent) qui apparaissent à plusieurs reprises. Cela permet aussi d'observer si certains facteurs ne s'observent qu'après un certain temps (avec ou sans 2023) ou que sur les parcelles avec de fortes ou faibles baseline.

- Pour le cas *Sans 2023*, cela correspond à toutes les parcelles concordantes de 2021 et 2022. Le choix de la suppression des parcelles de 2023 est lié à la faible croissance des parcelles de 2023 (peu de recul).
- Les cas *Uniquement baseline "faible"/"forte"* correspondent à toutes les parcelles concordantes (2023 incluses) ayant une baseline "faible" ou "forte". Les cas *Sans 2023, Uniquement baseline "faible"/"forte"* correspondent à toutes les parcelles concordantes (2023 exclues) ayant une baseline "faible" ou "forte".

  Nous faisons l'hypothèse que les facteurs influençant la croissance dépendent du type de baseline. Par exemple, nous pensons que le feu n'aura pas le même impact pour la baseline faible ou forte.

Nous avons aussi choisi de garder uniquement les facteurs ayant une influence significative (p-value inférieur à 0.05). Associées à ces facteurs, nous reportons en plus de la p-value la direction de l'influence du facteur, ainsi s'il est positivement corrélé à notre classification (fortes valeurs associées aux parcelles "bien" et faibles valeurs associées aux parcelles "pas bien"/"mauvaise") alors il apparaît en vert et en rouge s'il est négativement corrélé.

4 Résultats

4.1 Comparaison des méthodes de clustering

Nous allons commencer par comparer la non-concordance totale entre les deux méthodes dans le cas d'un clustering en deux ou trois clusters. La non-concordance représente une situation où les deux méthodes fournissent des clusters différents. Par exemple, si une parcelle est affectée au cluster "mauvais" par la méthode 1 et "moyen" par la méthode 2, cette parcelle représentera un cas de non-concordance.
En choisissant 2 clusters, la non-concordance totale entre les deux modèles est de 18.18%. D'un autre côte le clustering en 3 groupes génère une non-concordance de 23.75%. Ce qui correspond à une concordance de 81.82% et 76.25% respectivement. La table ci-dessous illustre les pourcentages de correspondance pour chaque catégorie entre les deux méthodes sur l'ensemble des 616 parcelles (Table 2).

Méthode 1 \ Méthode 2		2 clusters		3 clusters		
		Bien	Pas bien	Bien	Moyen	Mauvais
2 clusters	Bien	48.32%	13.15%			
	Pas Bien	5.03%	33.5%			
3 clusters	Bien			21.2%	1.24%	
	Moyen			9.74%	42.05%	2.37%
	Mauvais			0.66%	9.74%	13%

Table 2: Correspondances et différences entre les deux méthodes dans le cas d'un clustering en 2 clusters et en 3 clusters

Nous obtenons ainsi deux types de clustering : le premier divise les données en deux groupes, "bien" et "pas bien", tandis que le second, plus détaillé, répartit les données en trois groupes. Le second clustering permet de distinguer plus précisément deux groupes extrêmes, correspondant respectivement à une croissance très bonne ou mauvaise. Ces deux clustering seront utiles pour l'analyse des facteurs: le premier, dont les deux classes comportent un grand nombre de parcelles assez hétérogènes (les parcelles "moyennes" étant incluses dans le groupe "bien" respectivement "pas bien") et un second, avec moins de parcelles, mais dont les deux classes "bien" et "mauvais" sont plus homogènes et bien différenciées.

4.2 Facteurs anthropiques et naturels

4.2.1 Corrélations entre les facteurs

Avant d'analyser les relations entre les facteurs anthropiques et naturels et la croissance des parcelles, nous allons examiner la corrélation entre les différents facteurs quantitatifs que nous avons sélectionnés. Une matrice de corrélation est présentée sur la figure 5. Nous allons commenter ici toutes les corrélations supérieures à 0.5 ou inférieure à -0.5.

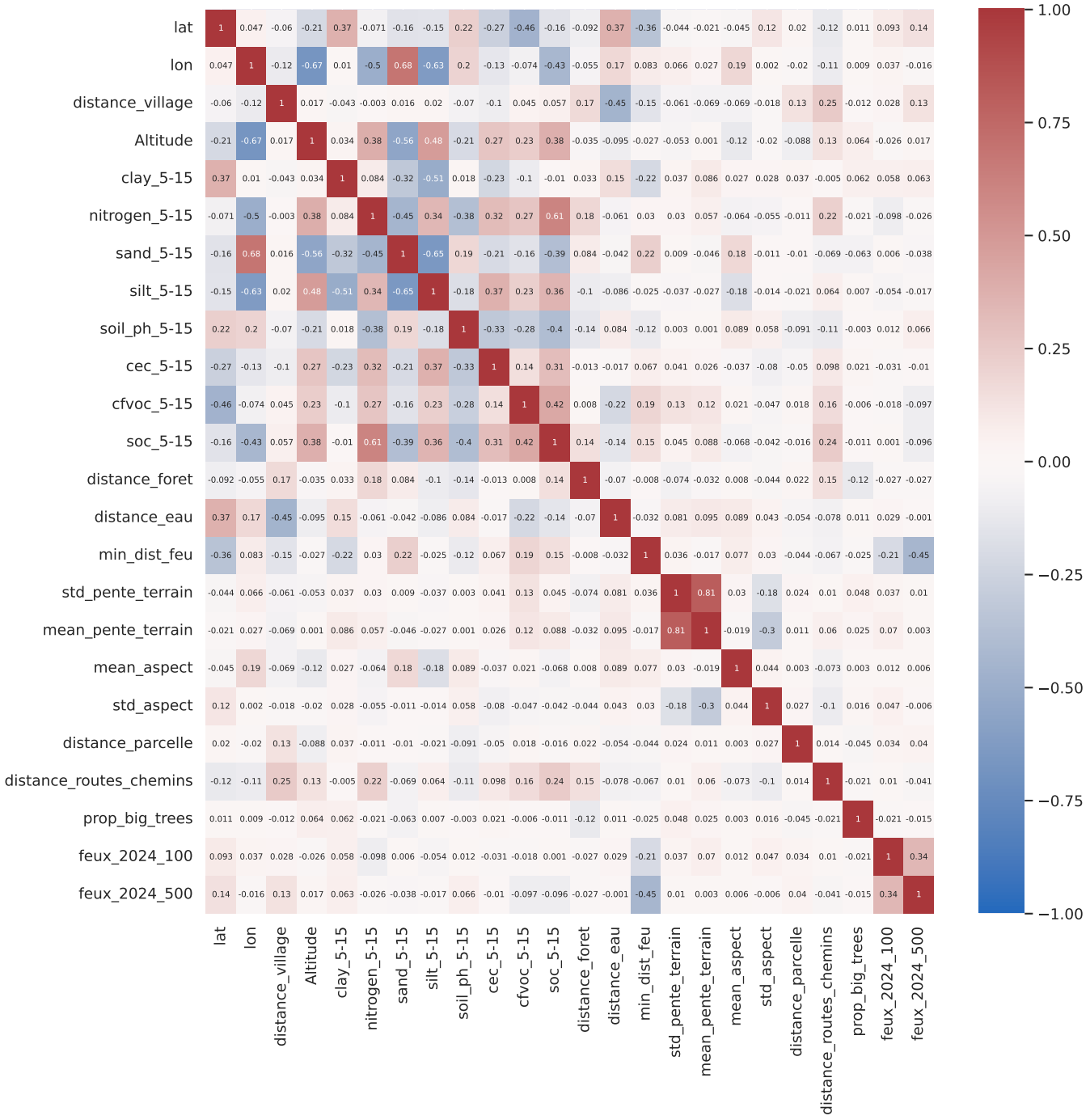

Figure 5: Matrice de corrélation entre les facteurs quantitatifs étudiés (représentation des coefficients de corrélation)

Nous remarquons en premier lieu une forte corrélation négative entre la longitude et l'altitude (*lon* et *Altitude*). Cette relation s'explique par la présence de la chaîne de collines du Simandou à l'ouest des parcelles, créant une région à plus haute altitude pour les faibles longitudes.
Nous remarquons également que plusieurs facteurs liés à la nature du sol sont corrélés entre eux, notamment une forte corrélation entre la quantité de sable et de limon (*sand_5-15* et *silt_5-15*). De plus, nous observons une corrélation positive entre les caractéristiques du sol et la longitude. En particulier, nous notons une corrélation moyenne positive entre la longitude et la concentration en sable (*lon* et *sand_5-15*). On retrouve donc les terrains plus sableux à l'Est, et les terrains plus limoneux à l'Ouest proche de la chaine du Simandou.

La pente moyenne et l'écart-type de la pente (qui mesure la variabilité de la pente sur le terrain) sont deux grandeurs corrélées avec une corrélation de 0.81 (*mean_pente_terrain* et *std_pente_terrain*). Ainsi, une pente moyenne grande est liée à une plus grande variabilité au sein du terrain, tandis qu'une pente faible est associée à une plus grande homogénéité de la pente du terrain.

4.3 Influence de la baseline sur la croissance des parcelles - facteur qualitatif

En utilisant la méthode de clustering des baselines, nous obtenons trois types de baselines. La première catégorie, appelée "baseline forte", représente les parcelles avec des valeurs élevées de NDVI en 2021, soit avant le reboisement par arboRise. La "baseline faible" désigne les parcelles comportant des pixels ayant des NDVI faibles (en 2021), tandis que la "baseline moyenne" correspond aux parcelles situées entre ces deux extrêmes. Nous obtenons 36.04% des parcelles ayant une baseline forte, 43.99% ayant une baseline moyenne et 19.97% ayant une baseline faible.
Par ailleurs, nous avons évalué l'écart-type des valeurs du NDVI à l'intérieur de chaque parcelle pour chaque catégorie. Nous avons constaté que les parcelles ayant une baseline forte présentent une déviation standard plus faible à l'intérieur de la parcelle, ce qui indique une uniformité plus élevée. Cela peut signifier que ces parcelles ont une baseline forte sur l'ensemble de la parcelle.
D'autre part, pour les parcelles de catégorie faible et moyenne, l'écart-type est en moyenne plus important. Ceci suggère que ces parcelles présentent une variabilité au sein de la même parcelle, avec une partie de la parcelle ayant une baseline forte et une autre partie une baseline faible (voir figure 6).

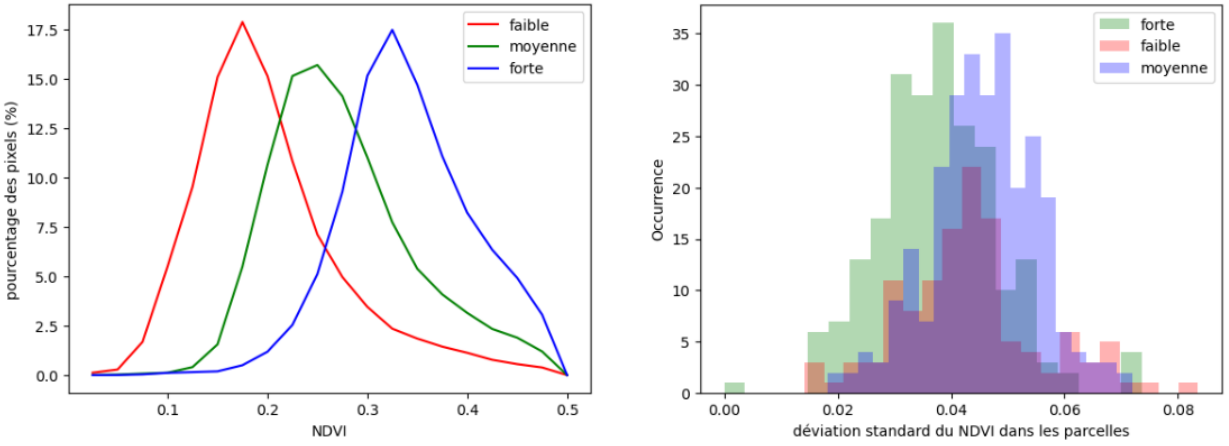

Figure 6: Répartition moyenne du NDVI sur les centroïdes des 3 catégories de baseline au 14/02/2021. Histogramme de la déviation standard du NDVI au 14/02/2021 des parcelles pour chaque catégories de baseline.

Nous avons évalué les proportions des parcelles "bien" et "pas bien" pour chaque type de baseline (forte, moyenne, faible), ainsi que la proportion des parcelles "bien", "moyen" et "mauvais" (Table 3).
Dans le cas de 2 clusters de croissance : pour la baseline "forte", nous observons à peu près autant de parcelles en croissance (bien) que de parcelles sans croissance (pas bien). Nous remarquons également que pour les baselines "faible" et "moyenne", la proportion de parcelles en croissance est plus importante que celle des parcelles sans croissance.
Dans le cas de trois clusters de croissance, pour une baseline "forte", le pourcentage de parcelles catégorisées comme "bien" est plus faible que celui des parcelles "moyennes" et "mauvaises" pour la méthode 1 de clustering de la croissance. Le ratio des parcelles "bonnes" par rapport aux parcelles "mauvaises" est également plus faible pour la méthode 2 dans le cas d'une baseline "forte", comparé à une baseline "moyenne" et "faible".

	Croissance	2 catégories				3 catégories					
		Méthode 1		Méthode 2		Méthode 1			Méthode 2		
		Bien	Pas bien	Bien	Pas bien	Bien	Moyen	Mauvais	Bien	Moyen	Mauvais
Baseline	Forte	49.1%	50.9%	42.34%	57.66%	17.56%	48.2%	34.24%	24.32%	54.95%	20.73%
	Moyenne	69%	31%	57.57%	42.43%	22.88%	61.62%	15.5%	34.68%	53.51%	10.81%
	Faible	66.67%	33.33%	64.23%	35.77%	29.26%	47.96%	22.78%	37.39%	47.96%	14.65%

Table 3: Pourcentage de parcelles dans chaque clusters pour les deux méthodes selon le type de baseline.

4.4 Relation entre les facteurs anthropiques et naturels et la classification de croissance des parcelles

L'analyse des facteurs anthropiques et naturels influençant les parcelles est réalisée en se basant sur les parcelles où les résultats des deux méthodes de classification sont concordants (Résultats 4.1).

Nous avons effectué l'analyse "Caractériser une variable qualitative" comme décrit dans la partie Méthodologie. Parmi tous les facteurs étudiés (Sources de données 2.2 & Table 1) ceux ressortant de cette analyse peuvent se diviser en 5 groupes.

Influence des facteurs sur la croissance (classification en 2 catégories)						
	Toutes les parcelles concordantes (n=502)	Sans 2023 (n=248)	Uniquement baseline "faible" (n=108)	Uniquement baseline "forte" (n=186)	Sans 2023, uniquement baseline "faible" (n=67)	Sans 2023, uniquement baseline "forte" (n=75)
clay_5-15	0.0121					
soc_5-15	0.0122	0.0241	0.0160		0.0153	
soil_ph_5-15	0.0135	0.0227		0.0484		
nitrogen_5-15	0.0231		0.0076		0.0132	
cfvoc_5-15	0.0297			0.0292		
mean_pente_terrain	0.0331					
min_dist_feu					0.0152	0.0318
feux_2024_500			0.0419			
prop_big_trees	0.0391					
Année	1.538e-12	4.805e-5	7.913e-3	1.264e-7		3.423e-5
Phase	8.651e-3	2.946e-4		3.582e-3		3.227e-4
Baseline	7.493e-3	0.042				
Village	8.134e-5					
Exposition				2.726e-3		

Table 4: Tableau des p-values associées aux facteurs étudiés en fonction du sous-jeu de données utilisé (n : nombre de parcelles), dans le cas de la classification en 2 clusters. La couleur (verte/rouge) représente respectivement une association positive/négative avec la classification des parcelles (positive signifiant que les fortes valeurs sont associées aux parcelles "bien" et les valeurs faibles aux parcelles "pas bien"). Pour les facteurs qualitatifs en bas du tableau, seul la p-value est reporté puisque la variable n'est pas quantitative. Les cases sont grisées lorsque il n'y a pas lieu de calculer une p-value (une seule classe présente par définition)

- Il y a d'abord l'influence de **la nature du sol**. On retrouve ici les facteurs reportant la composition du sol (*clay_5-15*, *soc_5-15*, *soil_ph_5-15*, *nitrogen_5-15*, *cfvoc_5-15*, *sand_5-15*), parmi eux les facteurs *sand_5-15* et *soil_ph_5-15* sont positivement corrélés à notre classification (Tables 4 & 5). A l'inverse les facteurs *clay_5-15*, *soc_5-15*, *nitrogen_5-15* et *cfvoc_5-15* sont négativement corrélés à notre classification (Tables 4 & 5). Cela signifie donc que les parcelles ayant de valeurs élevées pour ces facteurs sont moins propices à la croissance de la biomasse comme mesurée dans cette étude.

Influence des facteurs sur la croissance (classification en 3 catégories, "bien" et "mauvais" uniquement)						
	Toutes les parcelles concordantes (n=210)	Sans 2023 (n=108)	Uniquement baseline "faible" (n=50)	Uniquement baseline "forte" (n=74)	Sans 2023, uniquement baseline "faible" (n=28)	Sans 2023, uniquement baseline "forte" (n=29)
clay_5-15		0.0383				
soil_ph_5-15	0.0137	0.0459				
sand_5-15	0.0302	0.0071				
mean_pente_terrain	0.0229	0.0315	0.0470		0.0184	
std_pente_terrain	0.0210		0.0022			
feux_2024_500			0.0306			
distance_eau				0.0092		
Année	3.093e-10	4.725e-5	4.783e-4	1.551e-3		0.018
Phase	1.455e-3	2.890e-5		0.0441		0.031
Baseline	0.014					
Village	4.795e-3					

Table 5: Tableau des p-values associées aux facteurs étudiés en fonction du sous-jeu de données utilisé (n : nombre de parcelles), dans le cas de la classification en 3 clusters (les parcelles "moyennes" ayant été retirées pour essayer de faire ressortir les facteurs les plus importants). La couleur (verte/rouge) représente respectivement une association positive/négative avec la classification des parcelles (positive signifiant que les fortes valeurs sont associées aux parcelles "bien" et les valeurs faibles aux parcelles "mauvais"). Pour les facteurs qualitatifs en bas de la tableau, seul la p-value est reporté puisque la variable n'est pas quantitative. Les cases sont grisées lorsque il n'y a pas lieu de calculer une p-value (une seule classe présente par définition)

- **l'influence de la topographie:** On peut aussi ajouter l'influence de la pente du terrain (*mean_pente_terrain* et *std_pente_terrain*), ces 2 facteurs sont négativement corrélés avec notre classification et donc avec la croissance de la biomasse (Tables 4 & 5).

  Un autre facteur, l'exposition (*Exposition*) est à rattacher à ce groupe, c'est un facteur qualitatif qui correspond à la traduction de l'angle d'exposition de la parcelle en 8 points cardinaux (N, NE, E, SE, S, SW, W, NW). Il est significatif une seule fois pour les parcelles à baseline faible (incluant 2023) pour la classification en 2 groupes. Les parcelles exposées au Nord sont plutôt à croissance faible comparées aux parcelles exposées Sud dans ce cas.

- **Facteurs de proximité:** Nous pouvons également inclure le facteur *distance_eau*, qui apparaît une fois dans cette analyse (Table 5). Il s'agit d'un autre facteur lié à la position géographique de la parcelle, dans ce cas par rapport aux cours d'eau. À noter que le nombre de cours d'eau dans la région est assez faible, avec seulement trois cours d'eau principaux comme le Dion, un affluent du fleuve Niger.

- Un autre groupe peut être rattacher à **l'influence du feu**, deux facteurs (*min_dist_feu* qui prend en compte tous les feux de 2021 à 2024 et *feux_2024_500* spécifique à l'année 2024) sont à relier aux événement de feux. Ils sont tous les deux négativement corrélés à notre classification et donc à la croissance (Tables 4 & 5).

- Un dernier groupe est celui de l'année de plantation (*Année*), du groupe de plantation (*Phase*) des villages (*Village*) et de la baseline (*Baseline* & *prop_big_trees*). Ces facteurs ressortent aussi de ces analyses sans pouvoir leur associer d'effet positif ou négatif comme pour les variables qualitatives (sauf pour le facteur *prop_big_trees* qui a un effet négatif).

  Néanmoins il ressort que les parcelles de 2021 (Groupe pilote) performent moins bien, tout comme celles de 2023 par rapport à celles de 2022.

  Les parcelles de la Phase 3 (Groupe 3) semblent performer le mieux parmi les parcelles de 2022 et 2023, à l'inverse des parcelles de la Phase 2.

De même les parcelles à baseline "faible" performent mieux que celles à baseline "forte". On remarque aussi que dans un cas, la proportion d'arbres matures (la "Baseline en arbres matures") est négativement associée à la croissance des parcelles.

On observe aussi un effet des villages sur la croissance, certains d'entre eux ayant plus de parcelles à forte croissance que d'autres (voir la partie 5.3).

Il est intéressant de noter que les facteurs ressortant de notre analyse ne sont pas les mêmes en fonction de la baseline de la parcelle (voir section 5.2.1)

5 Discussion

5.1 Analyse d'un cas de non-concordance

Dans cette partie, nous allons nous focaliser sur un cas de non-concordance totale correspondant à un clustering en "mauvais" pour la méthode 1, indiquant une absence de croissance, et à un clustering en "bien" pour la méthode 2, indiquant une croissance (voir table 2). cette non-concordance a lieu pour les parcelles '2023-038', '2023-139', '2023-152', '2023-248'. Nous avons choisi la parcelle '2023-038' comme exemple pour étudier ce cas limite. La figure 10 en annexe représente une image satellite de la parcelles et des environs pour la date du 21/01/2024 ainsi que les valeurs du NDVI pour le 14/02/2021 sur la même zone.

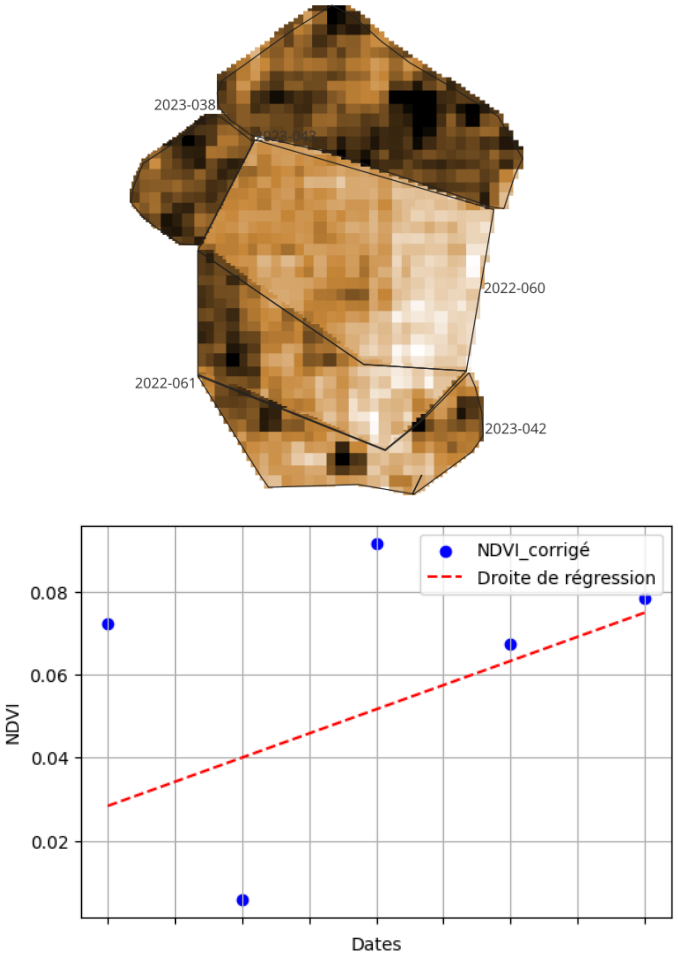

Figure 7: Valeur de la pente de la régression linéaire pour la parcelle 2023-038" et les parcelles adjacentes, les pixels les plus clairs correspondent à la plus forte pente; et pente de régression pour la parcelle 2023-038" obtenu avec la seconde méthode

Nous remarquons que cette parcelle est adjacente à d'autres parcelles dont la "2022-060". Puis un peu plus loin, la "2022-061" et la "2023-042". Les deux parcelles centrales (2022-060 et 2022-061), plantées toutes les deux en 2022, présentent la plus forte croissance (là où on voit le plus de pixels clairs, Figure 7). Les trois parcelles de 2023 affichent une croissance plus faible, probablement expliquée par un temps de croissance plus court, puisque ces parcelles n'ont eu qu'une seule année pour se développer. La méthode 2, qui prend en compte la différence par rapport aux valeurs du centre des villages, est censée mieux prendre en compte les variations naturelles du NDVI entre les dates de mesure. Peut-être que cela peut expliquer les différences qu'on observe sur la classification de ces

parcelles. Un autre aspect est le faible nombre de points pris en compte pour calculer la pente de régression pour les parcelles de 2023 (5 points). Ce faible nombre de points peut créer de l'incertitude dans les mesures. On observe sur le graphique de droite (Figure 7) que la deuxième valeur contribue fortement à rendre la pente calculée positive (et donc contribue fortement au classement de la parcelle dans la catégorie "bien" avec la deuxième méthode). Pour la méthode 1, une importante zone à l'est de la parcelle présente une faible croissance, résultant en un clustering qualifié de mauvais.
De manière générale, les quatre parcelles où l'on observe une non-concordance totale concernent uniquement des parcelles de 2023, avec peu de points et peu de temps de croissance. Cela nous amène à penser que notre méthode n'est pas toujours suffisamment sensible pour détecter correctement la croissance des parcelles de 2023, qui se trouvent encore au tout début de leur croissance.

5.2 Facteurs naturels et anthropiques influençant la croissance

5.2.1 Baseline

Nous avons vu dans la partie Résultats 4.3 que les parcelles en croissance sont plus souvent associées à une baseline faible qu'une baseline forte.
Nous pouvons émettre l'hypothèse qu'une baseline "forte", soit influence négativement la croissance de la parcelle, et cela même en retirant les arbres matures ayant une croissance faible/nulle du calcul de la croissance des parcelles, soit masque la croissance lié au reboisement par arboRise à cause d'une végétation buissonnante par exemple. De même pour les parcelles ayant une très faible baseline au départ, même une faible augmentation de biomasse va pouvoir très rapidement s'observer via le NDVI et ces parcelles ont ainsi plus de chance d'être dans les clusters de croissances "bien" que "mauvais" ou "pas bien". Dans tous les cas, la croissance de la biomasse des parcelles semble davantage liée au reboisement qu'à la croissance d'une quelconque baseline présente avant le reboisement.
On remarque aussi en lien avec les Tables 4 & 5 que la baseline a une influence sur les autres facteurs, les feux semblent ainsi avoir plus d'impacts sur les parcelles à baseline faible, Sûrement que ce sont des feux de faible intensité n'affectant que les petites plantes. De même le pH du sol a une influence plus importante sur les parcelles à baseline forte, et à l'inverse la pente semble avoir plus d'impact sur les parcelles à baseline faible. On peut supposer que la présence d'une baseline en amont du reboisement permet de mieux stabiliser le sol et d'éviter le ruissellement ce qui ne vas pas trop pénaliser la croissance.
La "baseline en arbres matures", différente de celle évoquée juste avant, est quantifiée par la surface relative qu'occupent les arbres matures sur les parcelles (*prop_big_trees*. Cette dernière est négativement associée à la croissance des parcelles. Notre classification des arbres matures n'étant pas parfaite, il se peut que des arbres matures restent non détectées et que leur croissance nulle tire vers le bas la croissance des parcelles. On peut aussi imaginer des effets de compétition et/ou d'ombrage sur les parcelles réduisant la croissance des nouveaux arbres.

5.2.2 Facteurs du mode de plantation et de gestion

Avec nos analyses, il est clair que l'année de plantation joue un rôle très important. L'année 2021 était plutôt mauvaise, l'année 2022 bonne et l'année 2023 moyenne. Deux effets sont probablement à l'oeuvre ici, un premier lié aux méthodes employées par arboRise, qui ont évoluées depuis le début du projet, et le second lié au temps nécessaire pour observer un effet de croissance pouvant expliquer que les parcelles de 2023 ne sont pas très bien représentées ici (et expliquant aussi notre choix d'exclure ces parcelles parfois).
De même le groupe de plantation (*Phase*), correspondant à la période de l'année à laquelle les arbres ont été plantés et aux espèces d'arbre plantées, a aussi un rôle dans la croissance. Il semble qu'il y a une meilleure croissance des parcelles du Groupe 3, il reste difficile de savoir si cela est lié aux espèces plantées dans ce groupe où la période de plantation plus favorable (plus tardive donc intervenant lors de la saisons humide par exemple).
Le village gestionnaire semble lui aussi jouer un rôle plus ou moins prépondérant dans la bonne (ou mauvaise) croissance des parcelles. Mais il est là encore difficile de dégager si l'effet est dû uniquement à la position géographique du village (nature différente de sol, topographie, etc) ou réellement au mode de gestion (intensité et période d'utilisation du feu, rotation de culture, etc). Nos analyses préliminaires semblent en tout cas confirmer qu'une part des différences peut s'expliquer par les données des facteurs abiotiques (sol, altitude, pente, etc) (Figure 8).
La distance aux autres parcelles (*distance_parcelle*), ainsi que la distance à la forêt la plus proche (*distance_foret*) ne semble pas avoir d'influence ici. Ces deux facteurs ne sont en tout cas pas ressortis de notre analyse. Peut-être que de meilleurs facteurs tels qu'un indice d'isolement des parcelles auraient pu être étudiés. Néanmoins aucun

schéma ne nous ait apparu en se penchant sur la carte de répartition des parcelles en fonction de leur classification (Figure ??, en Annexe).
La distance à l'eau (*distance_eau*) est négativement corrélée à la croissance, plus les parcelles sont loin de l'eau meilleur est leur croissance. Cette observation est contre-intuitive, cependant ce facteur n'apparaît qu'une seule fois dans nos analyses dans le cas des parcelles à baseline forte de 2021 à 2023 et uniquement pour le clustering en 3 groupes. Il se peut que cela ne soit qu'une erreur statistique. Néanmoins il se peut aussi que les parcelles très proches des cours d'eau se trouvent en fait dans les vallées avec donc une forte pente du terrain qui est défavorable à la croissance. Des événements de crues peuvent aussi avoir affecté négativement la croissance. Cela reste malgré tout un facteur dont l'influence est plutôt contre intuitive.

5.2.3 Facteurs liés aux perturbations naturelles ou anthropiques

Les feux jouent, semble-t-il, un rôle dans la bonne ou mauvaise croissance des parcelles. Le nombre d'événements de feux ayant eu lieu dans un rayon de 500m autour de la parcelle en 2024 (*feux_2024_500*), ainsi que la distance au feu le plus proche ayant eu lieu entre la période plantation et aujourd'hui (*min_dist_feu*) sont négativement corrélés à la bonne croissance des parcelles. Pour le nombre de feux en 2024 à moins de 500 mètres de la parcelle, cette corrélation semble cohérente avec nos attentes. En effet, si un feu a touché la parcelle, on peut s'attendre à une diminution du NDVI et donc de la croissance estimée, en lien avec la combustion de tout ou partie de la biomasse végétale. Pour les feux ayant pu avoir lieu entre 2021 et 2023, ils ont pu avoir un effet temporaire sur la croissance qui a pu reprendre après le feu avec la régénération naturelle, rendant ainsi plus difficile l'estimation de leurs effets avec notre méthode. La figure 11 en annexe représente une parcelle ayant probablement subi un feu le 4 février 2022, ce qui se manifeste par une chute du NDVI entre le 4 février et le 9 février. On remarque aussi la reprise de la croissance de la parcelle par la suite. Selon MODIS, un feu se situe à 66 mètres de la parcelle. Par ailleurs, la combustion de la végétation herbacée peut aussi favoriser la croissance en apportant des nutriments.
De plus nous ne disposons que de données partielles concernant les feux. Les données de MODIS ne reportent pas précisément l'étendue des feux mais seulement leur position à certains moments. Nous n'avons pas non plus accès à l'intensité de ces feux et donc à leur potentiel impact sur la végétation.
Enfin la distance aux feux mesurée (que cela soit pour calculer la distance au feu le plus proche ou le nombre de feux dans un certain rayon) ne prend pas en compte les spécificités du terrain (présence d'une barrière entre le feu et la parcelle comme une route par exemple). Pour mieux évaluer l'impact des feux, une connaissance plus précise de leur régime et de leur emplacement est donc nécessaire.

la distance aux routes et aux villages n'a pas semblé avoir d'influence particulière sur la croissance des parcelles. Néanmoins ne disposant pas de cartographie précise du réseau de routes et de chemins de la zone étudié, nous avons utilisé les images satellites pour tracer à la main le réseau de routes et chemin principaux (plus de 1m de large). Cette méthode est peu précise et peut être sujette à interprétation. De plus il se peut que certains chemins peu empruntés n'aient pas été pris en compte ou inversement. Le lien entre la position des routes et la propagation des feux (rôle de barrière mais aussi de réseau de propagation du feu) n'a pas été évalué bien qu'il puisse être crucial. Le lien entre présence de route et l'abroutissement n'est pas direct non plus. Nous pensons que les routes et chemin peuvent être utilisés par les troupeaux et animaux pour se déplacer et accéder aux parcelles, mais la topographie doit aussi être prise en compte.

5.2.4 Facteurs abiotiques

La nature du sol ressort en tête parmi les facteurs de notre analyse. Il semble qu'un sol plutôt à pH élevé (*soil_ph_5-15*) (moins acide) est plus propice à la croissance de la biomasse comme mesurée dans cette étude [pH,]. De plus, une forte teneur en sable (*sand_5-15*), faible teneur en argile (*clay_5-15*), faible teneur en matière organique (*soc_5-15*), faible teneur en azote (*nitrogen_5-15*) et à faible densité en larges fragments (*cfvoc_5-15*) soit le plus propice à la croissance (Tables 4 & 5).
Les fortes corrélations entre certaines variables de sol rend l'interprétation difficile. Par exemple on observe une corrélation positive avec un coefficient de 0.61 entre la teneur en matière organique et en azote (7 Annexe). Ainsi il est difficile de savoir lorsque nous observons un effet de ces deux variables, si une des deux est présente uniquement parce qu'elle est très corrélée à l'autre, ou si elle a un effet additionnel. Il est surprenant d'observer qu'une forte teneur en azote est corrélée négativement à la croissance des parcelles. Des effets lié à la baseline peuvent, peut-être

expliqué cette observation surprenante, en effet une forte teneur en azote favorise peut-être la présence d'une forte baseline en amont qui influence elle même négativement la croissance. De même la teneur en azote est négativement corrélée au pH (Figure 5) qui influence lui positivement la croissance, mais il est là encore difficile de savoir laquelle des variables est la plus importante.

Notre analyse fait également ressortir à plusieurs reprises la pente du terrain (*mean_pente_terrain* & *std_pente_terrain*) comme facteur influençant négativement la croissance des parcelles. Les parcelles ayant une forte pente (et très souvent aussi une forte variabilité dans la pente, Figure 5) sont plutôt associées à une croissance faible. Il se peut que les contraintes liées à la pente (ruissellement ou difficulté des arbres à prendre racine par exemples) expliquent cette observation.
L'effet de l'exposition (*Exposition*) qui n'apparaît qu'une fois dans l'analyse semble assez clair. Les parcelles exposées au Nord ont en moyenne une croissance plus faible que les parcelles exposées au Sud, très certainement en lien avec l'ensoleillement plus important au Sud qu'au Nord favorisant la croissance.
Les autres facteurs évoqués plus tôt (Figure 4) comme la latitude et longitude ou encore l'altitude ne semblent pas avoir d'impact direct sur la bonne ou mauvaise croissance des parcelles. Ils peuvent peut-être avoir des effets indirects en étant corrélés à d'autres facteurs. L'altitude est ainsi fortement corrélée à la longitude (corrélation négative ici, l'altitude diminue en allant vers l'Est) qui est elle même assez fortement corrélée à la teneur en sable ou en matière organique (Figure 5).

5.3 Analyse des villages gestionnaires propices et peu propices

Nous allons, dans cette partie, nous focaliser sur certains villages gestionnaires des parcelles afin d'évaluer les caractéristiques communes des "villages propices" au bon développement des parcelles et des "villages peu propices" à cela. Cette partie permet regrouper certaines parcelles et d'avoir une discussion macroscopique des facteurs d'influence.
Afin de déterminer les "villages propices" et les "villages peu propices", nous avons utilisé le clustering en trois clusters : "bien", "moyen" et "mauvais". Chaque village gère un certain nombre de parcelles, allant de 5 à 46. Pour chaque village, nous avons évalué le pourcentage de parcelles de cluster "bien" qui lui sont associées et le pourcentage de parcelles "mauvaises". Les villages propices sont ceux ayant plus de 40 % de bonnes parcelles, tandis que les villages peu propices sont ceux associés à plus de 35 % de mauvaises parcelles. Ces pourcentages correspondent aux valeurs au-dessus (ou en-dessous) de la moyenne plus (moins) l'écart type. Nous obtenons dans ce cas 6 "villages propices" et 5 "villages peu propices". Pour chacun des villages étudiés, nous avons estimé les valeurs moyennes des caractéristiques anthropiques et naturelles sur toutes les parcelles qui leur sont associées.
La figure 8 représente un box plot de plusieurs facteurs relatifs aux deux types de villages. On y retrouve des caractéristiques relatives au sol, telles que la concentration en azote, la concentration de sable et la teneur du sol en matière organique, comme décrit précédemment. On retrouve ici que les villages propices ont une plus forte concentration de sable (*sand_5-15*), une plus faible teneur en matière organique (*soc_5-15*) ainsi qu'une plus faible concentration d'azote (*nitrogen_5-15*) ce qui correspond aux analyses précédemment réalisées.

Par ailleurs, il semble que les "villages peu propices" soient situés plus à l'Ouest que les "villages propices" ce qui correspond à des altitudes plus élevées ainsi qu'à un relief plus accidenté. Un autre motif émerge concernant la distance entre les villages et les parcelles : On remarque ici que les parcelles gérées par des villages classés comme "peu propices" sont plus éloignées de ces villages par rapport aux parcelles associées aux villages classés comme "propices".

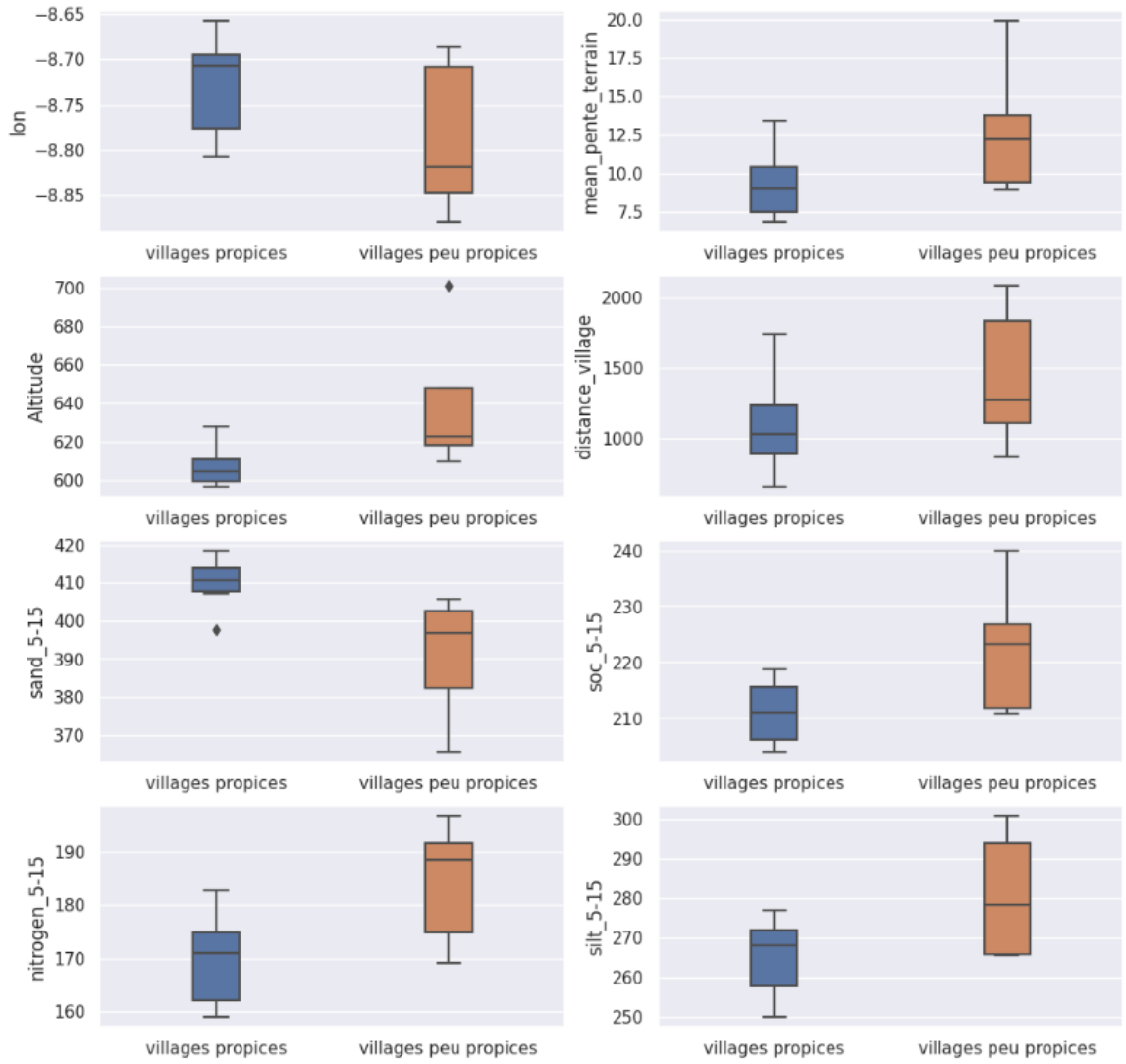

Figure 8: Caractéristiques des parcelles pour les villages propices et les villages peu propices .

5.4 Limites et perspectives

Limites

Notre méthode de classification se base uniquement sur des données satellitaires. Il conviendrait de coupler cela à des données de terrain pour pouvoir s'assurer réellement de comment une croissance du NDVI se traduit sur le terrain en terme de croissance de la biomasse ligneuse.

Nous avons choisi de retirer les pixels correspondant aux arbres dans notre analyse en fixant un seuil d'une valeur de NDVI de plus de 0.5 au 14/02/2021. Nous pourrions affiner ce critère en utilisant les images satellites pour identifier précisément les arbres matures et en entraînant par la suite un modèle de Random Forest pour reconnaître ces arbres.

Le calcul des distances pour les facteurs de proximité (distance route, distance village, etc) est basé sur la distance à vol d'oiseau entre les deux lieux. Les spécificités du terrain (topographie, obstacles, etc) ne sont pas prises en compte. Une amélioration potentielle serait de calculer les chemins de moindre résistance entre les variables d'intérêt.
Le nombre des parcelles pour la baseline faible (ou forte) présenté dans la table 5 est assez faible et peu limiter la

robustesse statistique de notre analyse.

Perspectives

Nous aurions aimé disposer de facteurs additionnels pour lesquels nous n'avons pas pu trouver de données avec une résolution spatiale suffisante (ou des données trop anciennes). C'est le cas notamment de l'humidité du sol dont nous aurions aimé connaître l'impact sur la croissance des parcelles. Mais malgré nos recherches aucune source de données avec une résolution spatiale inférieur à 500m n'a été trouvée.
Des facteurs météorologiques tels que la température et les précipitations n'ont pas été étudiés. En effet, la région ne dispose pas d'une couverture suffisante en stations météorologiques pour estimer les différences de précipitations sur la zone étudiée. De plus, nous pensons que ces facteurs météorologiques varient peu à l'échelle de la zone d'étude par rapport aux variations inter-annuelles, qui sont plus cruciales. Il est possible que des facteurs tels que la quantité de précipitations au moment de la plantation des graines (ou dans les semaines précédentes/suivantes) jouent un rôle dans la croissance des parcelles, mais nous ne disposions pas des données nécessaires pour mener à bien ces analyses.

6 Conclusion

Pour conclure, grâce à l'utilisation des données de Sentinel-2 et à l'utilisation de l'indice NDVI, nous avons pu estimer à partir de deux méthodes différentes la croissance de la biomasse ligneuse sur l'ensemble des parcelles arboRise.
Nos résultats révèlent que l'influence de la baseline est assez importante sur la croissance des parcelles et qu'elle doit être prise en compte par arboRise lors du choix des parcelles à reboiser. Les parcelles ayant une faible ou moyenne baseline sont celles où la croissance la plus forte a été observée et inversement pour les parcelles à forte baseline. Les autres facteurs influençant la croissance des parcelles sont liés à la composition du sol, des facteurs lié à la topographie de la parcelle ainsi que des facteurs liés aux événements de feux. Les fortes corrélations entre variables de sol rend leur interprétation difficile, néanmoins le pH du sol, sa teneur en argile, en azote et en matière organique semblent parmi les facteurs les plus importants. Le choix de l'emplacement de la parcelle est crucial et détermine en partie si la croissance y sera bonne ou mauvaise, néanmoins une certaine marge de manoeuvre est possible pour favoriser la croissance, comme limiter au maximum les feux d'atteindre les parcelles. Une attention particulière à la pente de la parcelle doit être prise en compte en tentant de favoriser les faibles pentes pour le choix des parcelles à reboiser. En revanche, la proximité des parcelles par rapport aux villages et aux routes ne semble pas avoir un effet statistiquement significatif sur la croissance.
Une étude plus approfondie avec des données sur le terrain ainsi que plus de recul (laisser plus de temps à la végétation pour pousser) serait nécessaire pour affiner et confirmer ces résultats préliminaires.

References

[pH,] Bibliothèque numérique DUDDAL · Les sols acides des tropiques · OMKS - BiblioSud.

[ALOS, 2020] ALOS, F.-F. (2020). Dataset | ALOS@EORC.

[AQUASTAT, 2021] AQUASTAT, F. (2021). AQUASTAT - FAO's global information system on water and agriculture.

[Askar et al., 2018] Askar, N. N., Phairuang, W., Wicaksono, P., and Sayektiningsih, T. (2018). Estimating aboveground biomass on private forest using sentinel-2 imagery. *Journal of Sensors*, 2018(16):6745629.

[Copernicus, 2021] Copernicus (2021). Copernicus digital elevation model.

[Dang et al., 2022] Dang, Y., Menzies, N., and Dalal, R. (2022). *Soil Constraints on Crop Production*. Cambridge Scholars Publishing.

[Foody et al., 2001] Foody, G. M., Cutler, M. E., McMorrow, J., Pelz, D., Tangki, H., Boyd, D. S., and Douglas, I. (2001). Mapping the biomass of bornean tropical rain forest from remotely sensed data. *Global Ecology and Biogeography*, 10(4):379–387.

[Kasekete, 2024] Kasekete, D. (2024). *Influence des facteurs biotiques et abiotiques sur la productivité forestière et la diversité floristique du sous-bois des plantations ligneuses autour du Parc National des Virunga en Province du Nord-Kivu (République Démocratique du Congo)*. PhD thesis.

[Koita, 1998] Koita, B. (1998). *Végétation post-culturale en zone soudanienne du Sénégal. Influence des pratiques culturales et des facteurs anthropiques sur la reconstitution végétale après abandon cultural*. These de doctorat, Corte.

[Lambert et al., 2015] Lambert, J., Denux, J.-P., Verbesselt, J., Balent, G., and Cheret, V. (2015). Detecting clear-cuts and decreases in forest vitality using modis ndvi time series. *Remote Sensing*, 7(4):3588–3612.

[Lang et al., 2022] Lang, N., Jetz, W., Schindler, K., and Wegner, J. D. (2022). A high-resolution canopy height model of the earth.

[Lebrun, 2022] Lebrun, J. (2022). Les apports de la télédétection pour le suivi de la végétation autour des ouvrages de restauration des terres au Sahel. Master's thesis.

[NASA, 2024] NASA (2024). NASA-FIRMS.

[Ouattara et al., 2022] Ouattara, B., Sanou, L., Koala, J., and Hien, M. (2022). Perceptions locales de la dégradation des ressources naturelles du corridor forestier de la Boucle du Mouhoun au Burkina Faso. *Bois et Forets des Tropiques*, 352:43–60.

[Potapov et al., 2021] Potapov, P., Li, X., Hernandez-Serna, A., Tyukavina, A., Hansen, M. C., Kommareddy, A., Pickens, A., Turubanova, S., Tang, H., Silva, C. E., Armston, J., Dubayah, R., Blair, J. B., and Hofton, M. (2021). Mapping global forest canopy height through integration of gedi and landsat data. *Remote Sensing of Environment*, 253:112165.

[Shi and Liu, 2017] Shi, L. and Liu, S. (2017). Methods of estimating forest biomass: A review.

[SoilGrids, 2024] SoilGrids (2024). SoilGrids — global gridded soil information.

[Vo Quang Truong, 2022] Vo Quang Truong, A. (2022). *Télédétection des forêts dégradées en Guinée à partir des données Sentinel-2*. Theses, Université Paris Cité.

7 Annexes

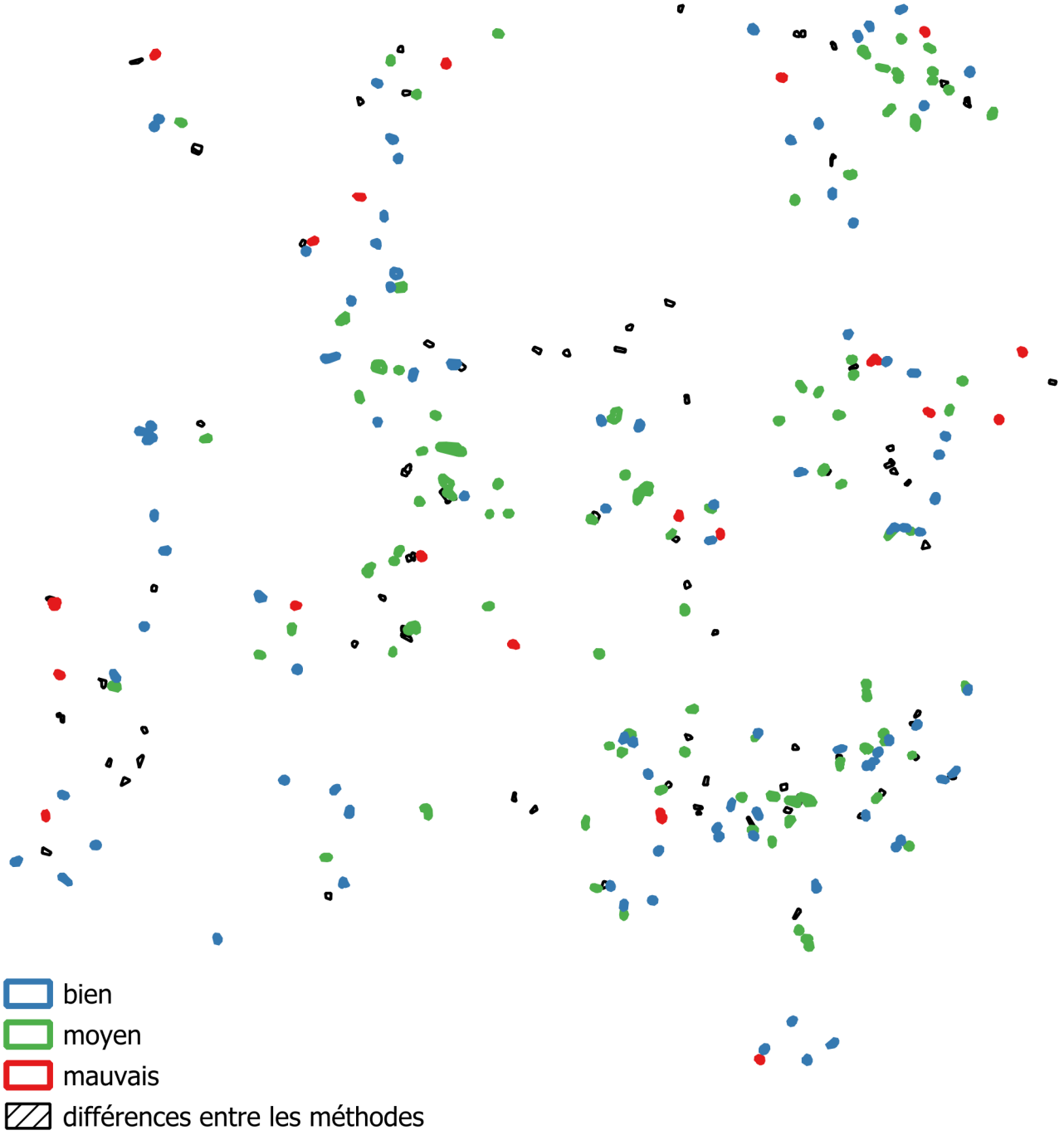

Figure 9: Localisation des parcelles classifiées en 3 catégories. Les catégories "bien", "moyen" et "mauvais" représentent les concordances de clustering entre les deux méthodes, la catégories "différences entre les méthodes" représente les parcelles avec non-concordance.

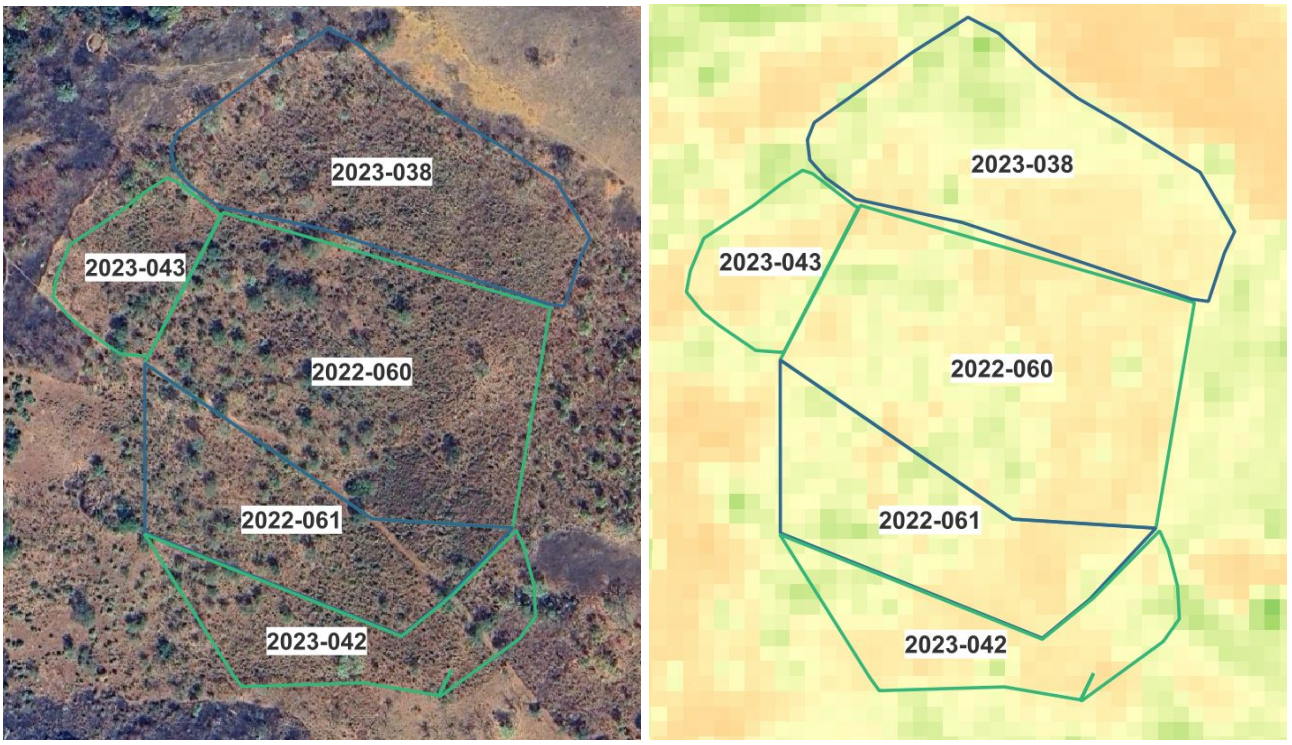

Figure 10: Image satellite de la parcelle "2023-038" et des environs (contenant les parcelles adjacentes) provenant de Google Satellite (en date du 21/01/2024). Et image satellite du NDVI obtenu à partir de Sentinel-2 au 14/02/2021 sur la même zone.

Année de reboisement	2021	2022	2023	2024
Date du passage de Sentinel-2 pris en compte dans l'étude	04/02/2021 09/02/2021 12/02/2021 24/02/2021	04/02/2022 09/02/2022 12/02/2022 19/02/2022	04/02/2023	09/02/2024 12/02/2024 19/02/2024 24/02/2024

Table 6: Date du passage de Sentinel-2 pris en compte dans l'étude

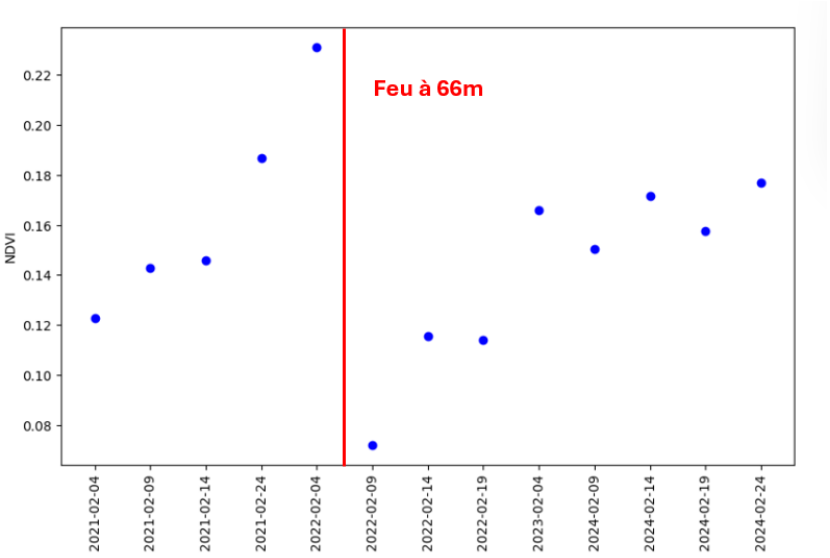

Figure 11: évolution du NDVI pour la parcelle 2022-050, un feu a eu lieu selon MODIS le 04-02-2022 à 66 mètres de la parcelle