

DESIGN PROJECT - RAPPORT FINAL

Conception d'une approche méthodologique pour la  
télédétection du leakage de carbone dans le cadre de  
projets de reforestation

*Application au projet ArboRise en Guinée Conakry*

Camille Dupré Tabti

Feryel El Phil

JUIN 2026

**EPFL**

SUPERVISEUR EPFL :  
CHRISTOPH BACHOFEN



SUPERVISEUR EXTERNE :  
PHILIPPE NICOD

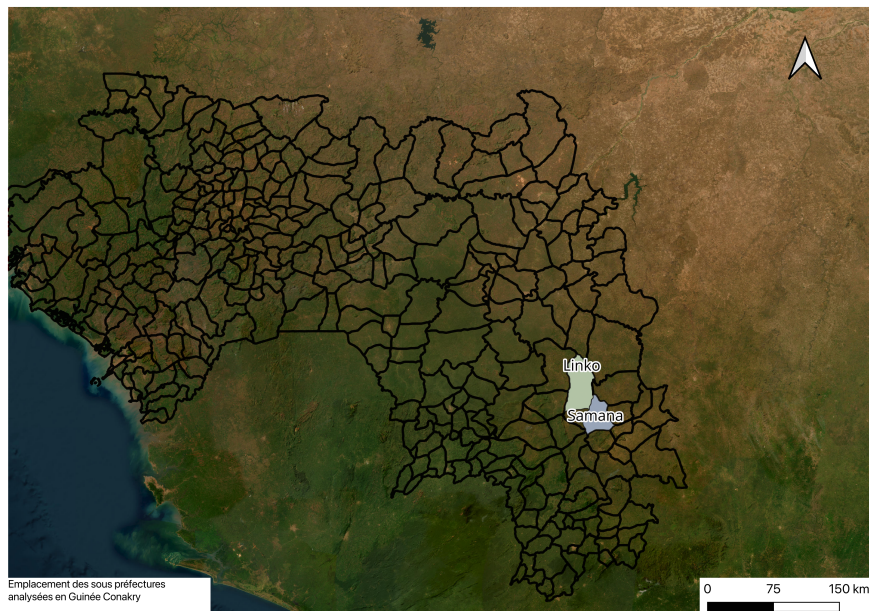
## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Données et méthodes</b>	<b>3</b>
2.1	Démarche et évolution méthodologique . . . . .	3
2.2	Sources de données . . . . .	3
2.3	Comparaison Hansen et NDVI : justification des choix de données . . . . .	4
2.4	Dispositif d'étude . . . . .	5
2.5	Reconstruction des cycles de jachère par méthode Tchebychev . . . . .	6
2.6	Détection du leakage par Difference-in-Differences . . . . .	7
2.7	Carte de risque de défrichement . . . . .	9
<b>3</b>	<b>Résultats</b>	<b>9</b>
3.1	Caractéristiques des parcelles . . . . .	9
3.2	Tendance régionale de perte de couverture arborée . . . . .	9
3.3	Analyse descriptive : comparaison Hansen et NDVI-dérivée . . . . .	10
3.4	Reconstruction des cycles de jachère par méthode Tchebychev . . . . .	12
3.5	Détection du leakage par Difference-in-Differences . . . . .	13
3.6	Carte de risque de défrichement . . . . .	13
<b>4</b>	<b>Discussion</b>	<b>15</b>
4.1	Reconstruction des cycles de jachère par méthode Tchebychev . . . . .	15
4.2	Détection du leakage par Difference-in-Differences . . . . .	15
4.3	Carte de risque de défrichement . . . . .	15
4.4	Limitations transversales, solutions et prochaines étapes . . . . .	16
<b>5</b>	<b>Conclusion</b>	<b>17</b>
	<b>Références</b>	<b>22</b>

# 1 Introduction

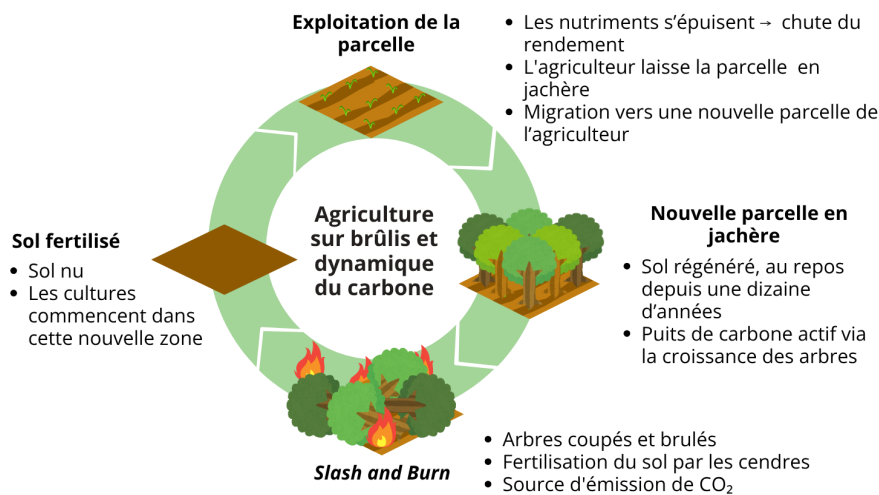
Dans le cadre de l'Accord de Paris, la reforestation est reconnue comme un outil important de séquestration du carbone. De nombreux pays ont intégré des objectifs de reboisement dans leurs contributions nationales, et des mécanismes de marché volontaire financent des projets portés par des organisations non gouvernementales et des entreprises privées, notamment en Afrique subsaharienne. Cependant, des études récentes ont mis en évidence que ces projets peuvent générer du *leakage* : un déplacement du défrichement vers d'autres zones, annulant partiellement ou totalement les bénéfices carbone comptabilisés. Les taux documentés de déplacement peuvent atteindre 12% de la surface reboisée dans un rayon de 150 km [1]. L'évaluation rigoureuse de ces effets nécessite des méthodes quasi-expérimentales robustes avec des groupes de contrôle explicitement définis, encore peu développées pour la plupart des contextes.

Le projet ArboRise conduit des activités de reforestation dans les sous-préfectures contiguës de Linko et de Samana en Guinée Conakry. Linko fait partie de la préfecture de Kérouané en Haute-Guinée, tandis que Samana appartient à la préfecture de Beyla en Guinée Forestière (Figure 1.1). Son modèle opérationnel repose sur un partenariat direct avec les familles rurales : certaines fournissent des graines d'essences locales, tandis que d'autres mettent à disposition certaines de leurs parcelles agricoles pour y conduire les plantations dans le cadre d'un mécanisme de compensation carbone.



**Figure 1.1:** Carte de situation géographique de la zone d'étude

La zone d'intervention est caractérisée par une agriculture itinérante sur brûlis. Les agriculteurs suivent un cycle bien défini (Figure 1.2) : les parcelles sont défrichées par *slash and burn* (coupe et brûlis), cultivées pendant cinq à huit ans jusqu'à épuisement de la fertilité du sol, puis laissées en jachère pendant huit à douze ans pour permettre la régénération naturelle. Dans ce contexte, il est essentiel de distinguer le *défrichement* (coupe d'arbres et brûlis sur une parcelle agricole en jachère, partie intégrante du cycle agronomique) de la *déforestation* (conversion permanente d'une forêt non agricole). Cette distinction définit précisément ce qu'on entend par *leakage* dans le cadre d'ArboRise.



**Figure 1.2:** Cycle classique de l'agriculture itinérante sur brûlis

Le projet ArboRise cible principalement les parcelles en fin de cycle cultural, c'est-à-dire lorsque la fertilité du sol est épuisée et que l'agriculteur laisserait normalement la parcelle en jachère. Au lieu de cette jachère naturelle, ArboRise y plante des essences sélectionnées pour leur potentiel de rendement (fruits alimentaires, usages médicaux ...). Ces arbres mettent cependant plus de temps à produire des rendements commercialisables qu'une culture agricole classique. Pendant cette période d'attente, l'agriculteur ne peut ni cultiver ni exploiter la parcelle pour générer un revenu. Pour compenser cette perte temporaire de ressource, il peut être amené à continuer ses activités agricoles habituelles sur ses autres parcelles. C'est donc durant cette phase d'attente que le risque de leakage émerge : l'agriculteur pourrait avoir besoin de défricher une autre de ses parcelles (en jachère) qu'il n'aurait pas défrichée sans le projet pour maintenir sa production agricole. Le leakage ne correspond donc pas à l'acte de défricher en lui-même, mais à une augmentation anormale du taux de défrichement causalement imputable à la contrainte du projet.

Bien que la mesure du leakage soit une obligation réglementaire, les méthodes actuelles restent moins développées dans le cadre des projets de reforestation et pas nécessairement adaptées à tous les contextes biophysiques et socio-économiques. ArboRise cherche donc à développer une méthodologie robuste et répliquable pour évaluer le leakage carbone dans son propre contexte opérationnel.

Un point méthodologique central conditionne cependant l'ensemble de cette étude. Selon le calendrier agronomique local, une parcelle plantée en 2021 ne serait naturellement disponible pour un nouveau cycle cultural qu'au bout de 8 à 12 ans : le leakage potentiel des premières cohortes ne serait donc observable qu'à partir de 2029 au minimum. Les données actuellement disponibles (2001–2024) ne capturent pas encore ce phénomène. Par conséquent, cette étude ne vise pas à mesurer le leakage réel, mais plutôt à concevoir une méthodologie robuste qui sera appliquée dès que le leakage deviendra observable.

L'étude poursuit trois objectifs : reconstruire les cycles de jachère historiques à partir de données satellitaires, établir une méthode de détection du leakage par Difference-in-Differences, et développer une carte de risque de défrichement futur pour un ciblage préventif.

## 2 Données et méthodes

### 2.1 Démarche et évolution méthodologique

La présente étude s'inscrit dans la continuité d'un projet antérieur conduit sur la même zone, qui avait tenté de mesurer les défrichements à partir d'un seuil absolu sur les valeurs de NDVI. Cette approche s'est révélée peu robuste : les valeurs absolues de NDVI varient fortement selon la latitude, la composition en espèces et les conditions climatiques locales, rendant tout seuil universel inadapté aux spécificités de chaque parcelle.

La présente étude a d'abord cherché à améliorer cette détection par NDVI en développant une méthode adaptative basée sur la dérivée première et l'inégalité de Tchebychev. Un défrichement est détecté par rapport aux variations antérieures de chaque parcelle, ce qui intègre automatiquement les conditions biophysiques locales sans paramètre absolu. Cette approche traite également une question posée par ArboRise : reconstruire les cycles de jachère historiques à partir de données satellitaires seules.

Cependant, pour la question centrale de mesure du leakage, le NDVI présente une limite fondamentale : il intègre l'ensemble du signal de végétation, incluant la variabilité saisonnière, les épisodes de sécheresse et les feux de surface, ce qui le rend difficile à interpréter de manière fiable à l'échelle régionale. Hansen Global Forest Watch, basé sur une classification supervisée multi-temporelle intégrant plusieurs indices spectraux, fournit une mesure plus directe et moins sensible aux artefacts saisonniers. Hansen a donc été retenu pour l'analyse Difference-in-Differences. À titre descriptif, une comparaison entre les deux approches est conduite pour illustrer leurs points d'accord et de désaccord (Section 2.3).

### 2.2 Sources de données

L'ensemble du traitement a été réalisé sur Google Earth Engine [2]. Le Tableau 2.1 récapitule les sources utilisées.

Source	Usage	Résolution	Période	Format
Hansen GFW v1.12	DiD, tendance régionale	30 m	2001–2024	Raster
Landsat 5/7/8/9	NDVI (Cycles de jachère et comparaison)	30 m	2000–2026	Raster
Sentinel-2 SR	Validation GPS	10 m	2015–2026	Raster
Polygones ArboRise	Unités d'analyse	–	2021–2026	Vecteur
JRC Tropical Moist Forest (TMF)	Dégradation de la couverture forestière	30 m	1982–2025	Raster
WorldPop	Densité de population	100 m	2015	Raster
Earthdata NASA	Modèle Numérique de Terrain (MNT)	30 m	2000	Raster
World Database on Protected Areas (WDPA)	Zones protégées	–	2026 (màj)	Vecteur
SoilGrids (ISRIC)	pH, Argile et CEC	250 m	v1.2 2014	Raster
Copernicus Global Land Cover Layers	Utilisation du sol (agricole vs forêt)	100 m	2015–2019	Raster

**Table 2.1:** Sources de données et leurs usages respectifs

**Hansen Global Forest Watch v1.12 [3].** Ce produit annuel de l'Université du Maryland cartographie la perte de couverture arborée à 30 mètres de résolution pour la période 2001–2024, à partir d'une classification supervisée pixel par pixel multi-temporelle intégrant plusieurs indices spectraux (NDVI, Normalized Burn Ratio (NBR), brightness, greenness, wetness). Un pixel est classé en perte si sa couverture arborée descend sous 10% à la suite d'une perturbation.

**Landsat 5, 7, 8 et 9.** Les quatre collections ont été fusionnées pour couvrir 2000–2026. Le masquage des nuages est effectué via la bande QA\_PIXEL (bits 3, 4 et 5). Pour chaque année, un composite médian est calculé sur la fenêtre novembre–mars (saison sèche), minimisant la couverture nuageuse et fixant le signal de végétation à son minimum physiologique annuel pour garantir la comparabilité inter-annuelle. Le NDVI est calculé selon le capteur :

$$\text{NDVI}_{L8/L9} = \frac{\text{SR\_B5} - \text{SR\_B4}}{\text{SR\_B5} + \text{SR\_B4}} \quad (1)$$

$$\text{NDVI}_{L5/L7} = \frac{\text{SR\_B4} - \text{SR\_B3}}{\text{SR\_B4} + \text{SR\_B3}} \quad (2)$$

**Sentinel-2.** Des composites mensuels Sentinel-2 SR ont été extraits sur juin 2015 – avril 2026. La résolution de 10 mètres avec une période de passage de 5 jours permettent d’observer la chute de NDVI au moment précis du défrichement, ce que le composite annuel Landsat ne peut résoudre avec un temps de passage de 16 jours.

### 2.3 Comparaison Hansen et NDVI : justification des choix de données

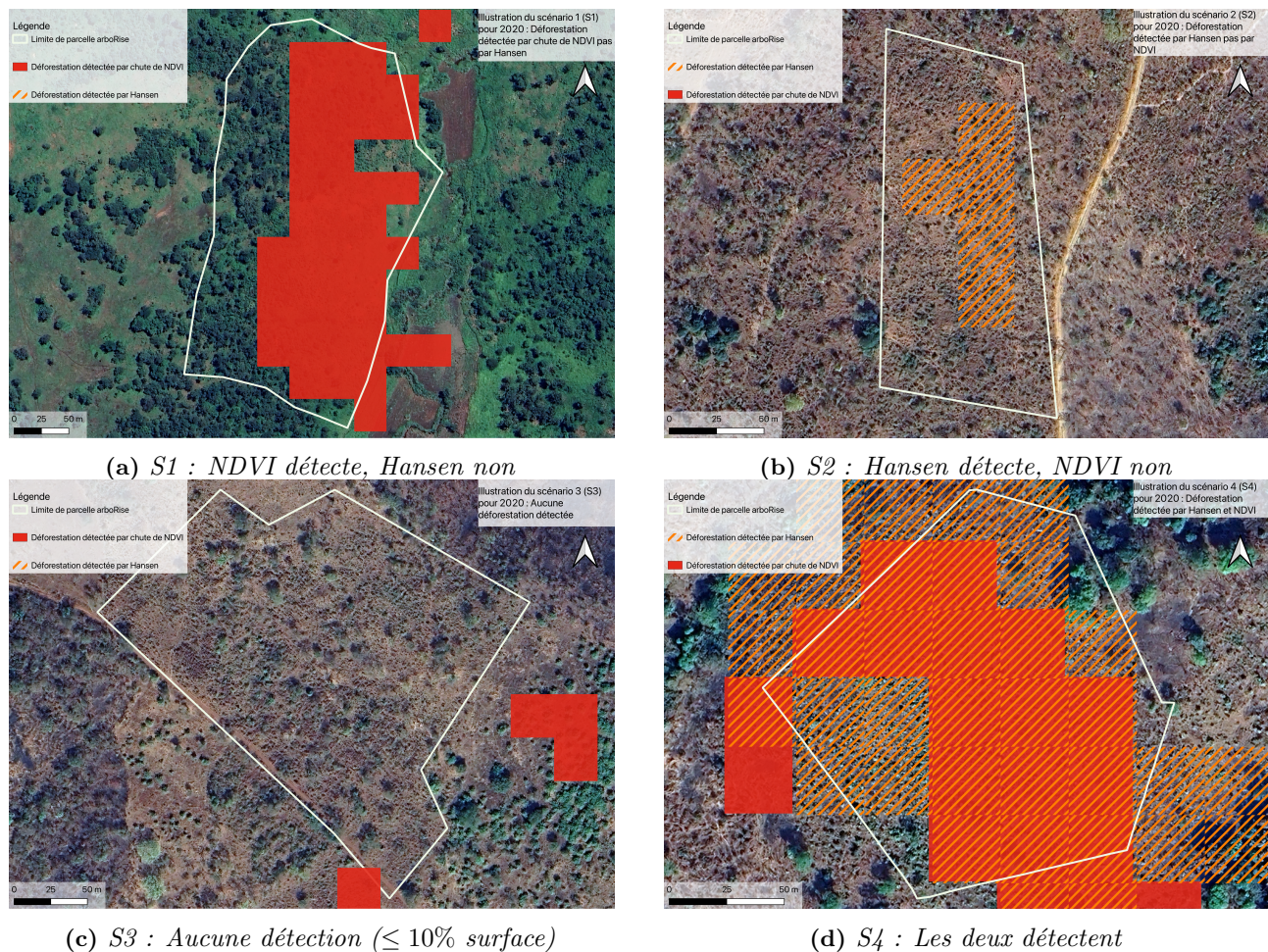
L’approche initiale de détection des défrichements reposait sur le NDVI seul. Le choix de Hansen pour l’analyse DiD résulte d’une différence structurelle : Hansen intègre plusieurs indices spectraux et offre une classification supervisée plus robuste qu’un simple indice de végétation. À titre descriptif, une comparaison entre les deux approches est conduite pour illustrer leurs points d’accord et de désaccord.

Pour chaque parcelle, deux indicateurs sont calculés indépendamment. D’une part, le taux de chevauchement entre les pixels Hansen et le polygone ArboRise : si ce taux dépasse 10% de la surface de la parcelle, Hansen est considéré comme détectant un défrichement. D’autre part, la dérivée du NDVI pixel par pixel entre deux années consécutives : les pixels dont la chute de NDVI dépasse le seuil calibré sur les points GPS de référence (Section 2.5) sont classés comme défrichés, et si ces pixels représentent plus de 10% de la surface de la parcelle, le NDVI est considéré comme détectant un défrichement.

Le Tableau 2.2 présente les quatre scénarios possibles issus de la combinaison de ces deux critères. La Figure 2.1 illustre chaque scénario sur des parcelles ArboRise réelles.

Scénario	Hansen	NDVI
S1	Non	Oui
S2	Oui	Non
S3	Non	Non
S4	Oui	Oui

**Table 2.2:** Scénarios de cohérence entre détections Hansen et NDVI. *Oui* : perte détectée (>10% de la surface de la parcelle). *Non* : pas de perte détectée.

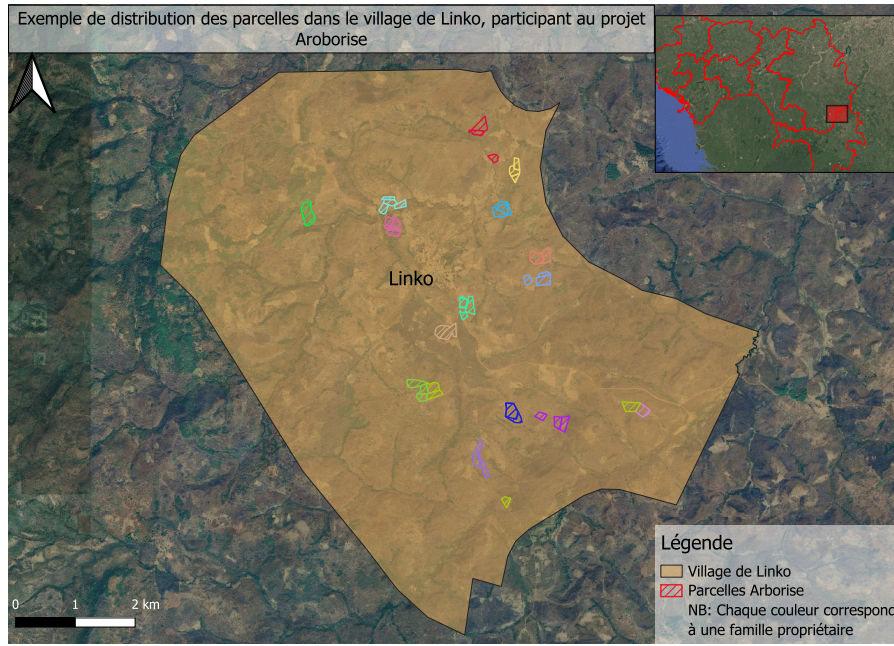


**Figure 2.1:** Illustration des quatre scénarios de cohérence entre Hansen et NDVI sur des parcelles ArboRise réelles pour l'année 2020

## 2.4 Dispositif d'étude

### 2.4.1 Unités d'analyse

Les parcelles ArboRise constituent les unités d'analyse : 840 polygones à Linko et 267 à Samana, chacun représentant une parcelle agricole familiale complète. La tendance régionale de perte de couverture arborée est calculée sur les limites administratives GADM des deux sous-préfectures. Toutes les autres analyses sont conduites directement sur les polygones individuels.



**Figure 2.2:** Exemple de distribution des parcelles ArboRise dans le village de Linko, sous-préfecture de Linko

## 2.5 Reconstruction des cycles de jachère par méthode Tchebychev

La méthode NDVI-Tchebychev a été développée pour améliorer la détection des défrichements par rapport à l’approche à seuil absolu du projet antérieur et reconstruire les cycles de jachère historiques à partir de données satellitaires seules.

Pour chaque parcelle, une série temporelle de NDVI annuel durant la saison sèche (novembre–mars) est extraite sur 2000–2026. Cette saison a été choisie pour minimiser la couverture nuageuse et fixer le signal de végétation à son minimum physiologique annuel, garantissant la comparabilité inter-annuelle.

Un défrichement génère une chute brutale du NDVI d’une année sur l’autre, visible comme un minimum local de la dérivée première. Pour tester si cette chute constitue une anomalie significative par rapport à la variabilité naturelle de la parcelle, on applique l’inégalité de Tchebychev, qui définit une borne supérieure sur la probabilité qu’une variable aléatoire s’écarte de sa moyenne au-delà d’un seuil donné, sans hypothèse sur la distribution sous-jacente [4].

Pour tout  $\epsilon > 0$ , l’inégalité de Bienaymé-Tchebychev s’écrit :

$$P(|X - E(X)| \geq \epsilon) \leq \frac{V(X)}{\epsilon^2} \quad (3)$$

En posant  $\epsilon = k\sigma$ , où  $\sigma = \sqrt{V(X)}$  représente l’écart-type de  $X$  et  $k > 0$ , l’inégalité se reformule ainsi :

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2} \quad (4)$$

Un défrichement est détecté en année  $t$  si :

$$\text{NDVI}(t-1) - \text{NDVI}(t) > k \cdot \sigma_i \quad (5)$$

$$t - t_{\text{dernier}} \geq n_{\text{min}} \text{ ans} \quad (6)$$

où  $\sigma_i$  est l’écart-type de la série NDVI de la parcelle  $i$  sur 2000–2026,  $k$  le paramètre de sensibilité et  $n_{\text{min}}$  la durée minimale entre deux défrichements consécutifs. Le seuil  $k \cdot \sigma_i$  est adaptatif à deux niveaux. D’une part, il est calculé individuellement pour chaque parcelle à partir de sa propre variabilité historique, sans recourir à un seuil absolu sur la valeur du NDVI. D’autre part, un défrichement est détecté par rapport aux variations antérieures de cette même parcelle. Cette propriété résout le

problème identifié dans l'étude antérieure, où un seuil absolu unique ne pouvait pas rendre compte des conditions biophysiques locales variables d'une parcelle à l'autre.

La méthode est validée sur trois points GPS de défrichements confirmés entre décembre 2025 et janvier 2026 : Sogboro ( $-8.513E, 9.027N$ ), Borohila ( $-8.803E, 9.392N$ ) et Koyola ( $-8.859E, 9.329N$ ).

Pour  $k = 1.5$  et  $n_{\min} = 8$  ans, l'inégalité de Tchebychev garantit qu'une chute dépassant  $1.5\sigma$  a moins de 44% de probabilité d'être due à la variabilité ordinaire.

L'analyse repose sur une approche multi-échelle complémentaire :

- *Landsat (interannuel, 2000–2026)* : Le profil annuel est comparé à sa moyenne historique 2000–2023. Exclure la période 2024–2026 de cette ligne de référence évite la contamination statistique de la moyenne et de la variance par les phases préparatoires et finales du défrichage.
- *Sentinel-2 (mensuel, 2015–2026)* : Grâce à sa haute fréquence temporelle, cette série permet de dater précisément l'événement au mois près et de distinguer la coupe rase d'une simple sécheresse saisonnière.

## 2.6 Détection du leakage par Difference-in-Differences

### 2.6.1 Choix des parcelles témoins

Pour mesurer un leakage induit par ArboRise, il faut comparer les parcelles engagées dans le projet à des terres agricoles similaires non soumises à la contrainte ArboRise pendant la période d'analyse.

Une première approche envisagée consistait à délimiter manuellement des parcelles témoins à partir d'images satellites, en identifiant visuellement les limites des terrains agricoles. Cette méthode s'est révélée trop peu fiable et trop chronophage pour obtenir un nombre de témoins statistiquement suffisant : les limites des parcelles agricoles sont difficiles à distinguer avec précision et la démarche ne serait pas répliquable à grande échelle.

La solution retenue consiste à utiliser comme témoins les parcelles ArboRise dont l'entrée dans le projet est postérieure à 2024. L'hypothèse sous-jacente est que ces terrains étaient déjà des terres agricoles actives en 2021. Comme notre période d'analyse couvre 2021–2024, ces parcelles n'ont subi aucune contrainte ArboRise pendant cette fenêtre et peuvent servir de contrefactuel. Cette approche présente l'avantage décisif de fournir des polygones précis et fiables, comparables en format aux polygones des parcelles traitées. Elle est également répliquable dans le futur : en décalant la limite entre parcelles traitées et parcelles témoins, la même méthode peut-être appliquée sans modification.

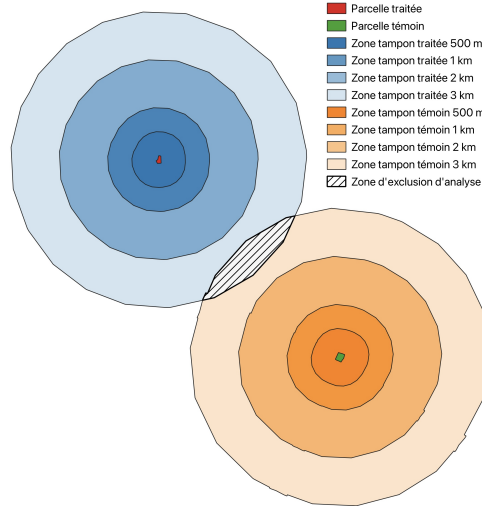
Le groupe témoin retenu comprend 195 parcelles : 145 parcelles Samana dont l'entrée dans le projet ArboRise date de 2025–2026 et 50 parcelles d'un projet ArboRise (biochar) dans la région de Linko dont l'entrée dans le projet date de 2026. Le groupe traité comprend 62 parcelles à Linko plantées en 2021, cohorte disposant du recul temporel le plus important (quatre ans jusqu'en 2024).

### 2.6.2 Zone de mesure du leakage potentiel

Le leakage ne se produit pas sur les parcelles ArboRise elles-mêmes mais sur d'autres parcelles familiales, qui se situent toutes dans le périmètre du village. Cependant, nous ne disposons que de la liste des parcelles engagées dans le projet ArboRise ; nous n'avons pas l'inventaire exhaustif de toutes les terres détenues par chaque agriculteur. Pour évaluer le leakage potentiel, on mesure donc la perte de couverture arborée dans un buffer circulaire autour de chaque parcelle, en considérant les cinq parcelles témoins les plus proches. La même approche est appliquée symétriquement aux parcelles témoins : si le buffer est un proxy imparfait, il l'est de la même façon pour les deux groupes, préservant la comparabilité du DiD.

Idéalement, on analyserait l'ensemble du périmètre du village. Cependant, les limites précises des villages pour Samana n'étaient pas disponibles. Pour se concentrer sur la méthodologie de détection du leakage elle-même plutôt que de consacrer du temps à la délimitation géographique, on a adopté la

simplification des buffers circulaires. Quatre rayons sont testés (500m, 1000m, 2000m, 3000m). Cette approche, bien qu'imparfaite, reste facilement modifiable dans le futur pour incorporer les limites de villages une fois disponibles.



**Figure 2.3:** Illustration des quatre tailles de buffer testées autour d'une parcelle traitée et d'une parcelle témoin. La parcelle témoin est l'une des 5 parcelles témoins situées dans le rayon de 10 km autour de la parcelle traitée.

### 2.6.3 Méthode Difference-in-Differences

La méthode DiD compare l'évolution de la perte de couverture arborée entre groupe traité et groupe témoin avant et après le démarrage du projet, contrôlant les tendances temporelles communes. Si le projet génère du leakage, la perte de couverture arborée devrait augmenter davantage autour des parcelles traitées qu'autour des témoins après le lancement du projet.

L'intensité de perte de couverture arborée annualisée dans un buffer de rayon  $d$  est :

$$I_{i,p} = \frac{D_{i,p}}{S_i \cdot n_p} \quad (7)$$

où  $D_{i,p}$  est la surface de perte de couverture arborée (Hansen) en  $\text{km}^2$  dans le buffer de la parcelle  $i$  pendant la période  $p$ ,  $S_i$  la surface du buffer en  $\text{km}^2$ , et  $n_p$  le nombre d'années ( $n_{\text{baseline}} = 20$  pour 2001–2020 ;  $n_{\text{suiivi}} = 4$  pour 2021–2024). La normalisation par  $n_p$  est indispensable : sans elle, l'intensité cumulée sur la baseline serait mécaniquement cinq fois plus grande que sur la période de suivi.

Le delta d'intensité annualisée est :

$$\Delta_i = I_{i,\text{suiivi}} - I_{i,\text{baseline}} \quad (8)$$

Pour chaque parcelle traitée  $i$ , les  $K = 5$  témoins les plus proches géographiquement sont sélectionnés dans l'anneau  $[d, 10 \text{ km}]$  par algorithme KD-Tree. La borne inférieure  $d$  constitue une zone d'exclusion stricte garantissant l'absence de chevauchement entre les buffers du traité et de ses témoins, éliminant le biais de contamination croisée. Le DiD individuel est :

$$\text{DiD}_i = \Delta_i^{\text{traité}} - \frac{1}{K} \sum_{j \in \mathcal{V}(i)} \Delta_j^{\text{témoin}} \quad (9)$$

Un  $\text{DiD}_i > 0$  indique un signal de leakage potentiel. Quatre rayons sont testés (500m, 1000m, 2000m, 3000m) ; un signal est considéré robuste s'il est cohérent à toutes les échelles. La significativité est évaluée par test de Wilcoxon ( $H_0$  : DiD médian = 0,  $\alpha = 0.05$ ) avec intervalles de confiance à 95% par bootstrap ( $n = 1000$ ).

## 2.7 Carte de risque de défrichement

La carte de risque est générée à l'aide d'un algorithme d'apprentissage supervisé de type Random Forest. Cette méthode combine les prédictions de plusieurs arbres de décision indépendants, entraînés sur des sous-ensembles aléatoires de données, afin d'obtenir une classification plus stable et robuste qu'un arbre unique.

Le modèle est alimenté par un ensemble de variables explicatives (détaillées dans le Tableau 2.1) incluant la densité de population, la proximité des zones protégées, les propriétés physico-chimiques du sol (pH, teneur en argile, capacité d'échange cationique), la pente ainsi que la distance au réseau routier. Pour l'apprentissage, les données historiques de perte de couverture arborée servent de vérité terrain, l'objectif du modèle étant de maximiser sa capacité prédictive vis-à-vis de ces événements de perturbation.

## 3 Résultats

### 3.1 Caractéristiques des parcelles

Le projet ArboRise couvre 1 157 parcelles réparties en trois groupes (Tableau 3.1). Les parcelles de Linko ( $n = 840$ , moyenne 2,17 ha) constituent le groupe où le projet est le plus développé, tandis que Samana présente des exploitations plus vastes ( $n = 267$ , moyenne 5,80 ha). Le groupe biochar ( $n = 50$ , moyenne 2,11 ha) présente une homogénéité comparable à Linko.

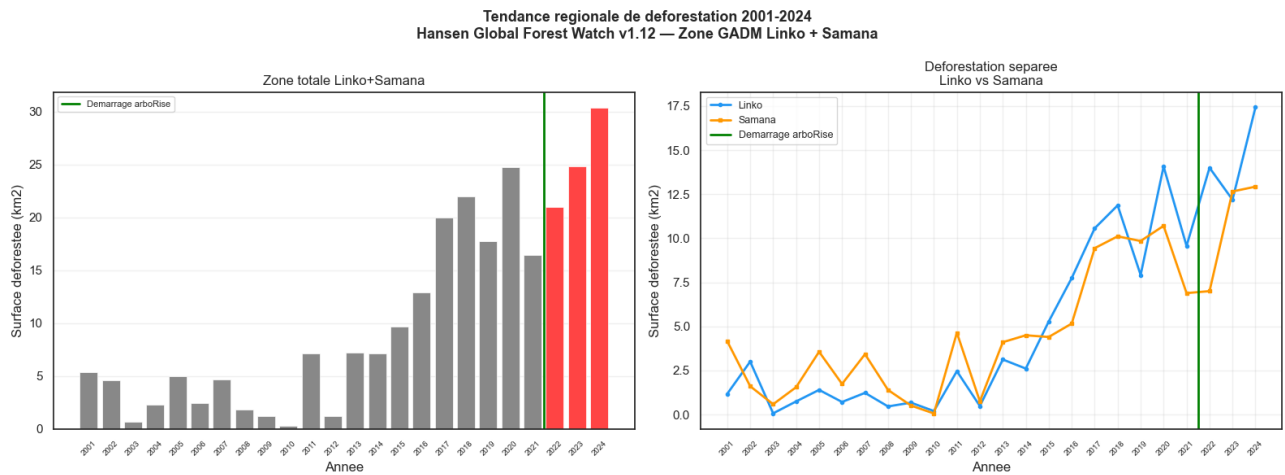
Groupe	n	Moyenne (ha)	Max (ha)	Min (ha)
Linko	840	2,17	9,50	0,10
Samana	267	5,80	8,09	0,44
Biochar (Linko)	50	2,11	3,56	0,99
<b>Total</b>	<b>1 157</b>	<b>3,36</b>	<b>9,50</b>	<b>0,10</b>

**Table 3.1:** *Statistiques descriptives de la superficie des parcelles ArboRise par groupe.*

### 3.2 Tendances régionale de perte de couverture arborée

L'objectif principal étant de mesurer le leakage de carbone, donc uniquement les défrichements agricoles induits par le projet ArboRise, nous analyserions idéalement uniquement les défrichements cycliques (agriculture itinérante), excluant la déforestation forestière qui ne relève pas de notre sujet. Cependant, Hansen mesure la perte de couverture arborée sans distinction entre défrichements agricoles et déforestation forestière. L'analyse à l'échelle de l'ensemble des deux sous-préfectures révèle une augmentation continue de cette perte à partir de 2014, antérieure au démarrage du projet ArboRise en 2021 (Figure 3.1). Cette tendance inclut donc à la fois les conversions forestières et les défrichements agricoles cycliques. Aucune des données accessibles actuellement ne permettent de les différencier.

Les deux sous-préfectures présentent des tendances régionales similaires, avec une accélération à partir de 2015 et un pic en 2024. Puisque cette tendance affecte pareillement les deux régions, elle impacte aussi bien les parcelles traitées que les témoins dans notre mesure du leakage.



**Figure 3.1:** Évolution annuelle de la perte de couverture arborée (Hansen GFW v1.12) sur 2001–2024 dans les sous-préfectures de Linko et Samana. La ligne verticale indique le démarrage du projet ArboRise en 2021.

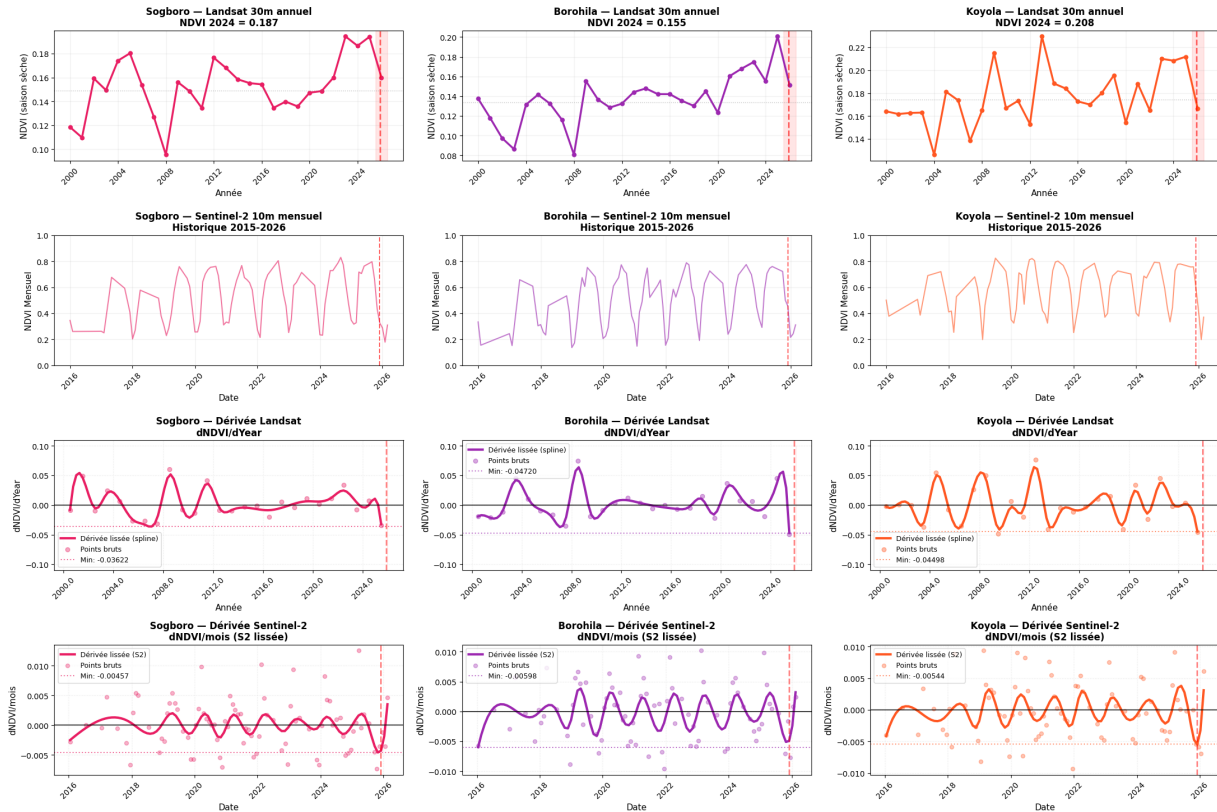
### 3.3 Analyse descriptive : comparaison Hansen et NDVI-dérivée

Pour évaluer les différences entre les deux approches de détection, une comparaison descriptive est conduite sur 2017–2021. Pour chaque année, deux indicateurs sont calculés : Hansen (seuil 10% de chevauchement pixel) et NDVI-dérivée (seuil 0.034 calibré sur des défrichements GPS confirmés). Cette comparaison vise à visualiser les deux approches plutôt qu'à déterminer laquelle est supérieure. Le choix de privilégier Hansen pour l'analyse du leakage repose sur sa couverture plus complète des données. Cette comparaison permet néanmoins d'illustrer la sensibilité inhérente à la méthode NDVI-dérivée.

#### 3.3.1 Validation terrain du seuil NDVI

Trois points de défrichement confirmés sur le terrain (décembre 2025 – janvier 2026) à Sogboro, Borohila et Koyola ont servi à valider la détection par dérivée NDVI. Le seuil de 0.034 correspond à la médiane des chutes NDVI observées à ces points. La Figure 3.2 montre les profils NDVI Landsat et Sentinel-2 sur ces trois sites. L'étude de la dérivée première s'avère pertinente pour mettre en évidence la chute brutale du NDVI, avec un minimum global apparaissant au moment du défrichement. Les données Landsat permettent d'identifier l'année de défrichement tandis que Sentinel-2, grâce à sa haute fréquence temporelle, permet de dater le mois de l'événement.

**Validation GPS — Analyse NDVI Complète avec Dérivées Lissées & Minima**  
**Ligne 1 : Landsat annuel (30m) saison sèche 2000-2026**  
**Ligne 2 : Sentinel-2 mensuel (10m) historique 2015-2026**  
**Ligne 3 : Dérivées Landsat — dNDVI/dYear par site avec ligne de minimum**  
**Ligne 4 : Dérivées Sentinel-2, dNDVI/mois par site avec ligne de minimum**



**Figure 3.2:** Profils NDVI et dérivée première (Landsat 30 m, composite saison sèche novembre–mars) sur les trois points de validation GPS, 2000–2026.

### 3.3.2 Résultats de la comparaison

Le choix du seuil de détection NDVI a un impact majeur sur les résultats. Le Tableau 3.2 illustre cette sensibilité en montrant comment le taux d'accord avec Hansen varie selon le seuil retenu pour l'année 2020.

Seuil NDVI	S1	S2	S3	S4	Accord Hansen-NDVI
0.010	334	22	772	29	56,9%
0.020	137	29	969	22	43,1%
0.030	31	40	1075	11	21,6%
0.034	17	42	1089	9	17,6%
0.050	3	49	1103	2	3,9%

**Table 3.2:** Sensibilité des résultats de comparaison Hansen/NDVI au choix du seuil de détection NDVI pour l'année 2020. La colonne Accord Hansen-NDVI indique le pourcentage de parcelles confirmées par les deux méthodes parmi les 51 parcelles où Hansen détecte une perte. Les résultats pour les cinq années 2017–2021 sont présentés en Annexe 5.1.

Pour le seuil calibré GPS (0.034), Hansen détecte 51 parcelles avec perte de couverture arborée en 2020 contre 26 pour la dérivée NDVI. La matrice de cohérence (Tableau 3.3) montre que parmi les 51 parcelles détectées par Hansen, 9 sont simultanément confirmées par le NDVI (S4). Les 42 parcelles restantes (S2) correspondent à des défrichements que Hansen détecte mais que le composite annuel NDVI ne résout pas.

	Hansen détecte	Hansen ne détecte pas
NDVI détecte	S4 : 9	S1 : 17
NDVI ne détecte pas	S2 : 42	S3 : 1089

**Table 3.3:** Matrice de cohérence Hansen/NDVI pour 2020 sur 1 157 parcelles. L'analyse inclut toutes les parcelles car, en l'absence de vérité terrain exhaustive, il est impossible de déterminer précisément lesquelles ont été défrichées en 2020. Cette couverture complète maximise la disponibilité de données pour la comparaison, expliquant le nombre élevé de parcelles où aucune des deux méthodes ne détecte de changement (S3)

### 3.4 Reconstruction des cycles de jachère par méthode Tchebychev

#### 3.4.1 Calibration de la méthode Tchebychev

Trois points de défrichement confirmés sur le terrain ont permis de calibrer le multiplicateur  $k$  de la méthode Tchebychev. Le Tableau 3.4 montre les paramètres de chaque site.

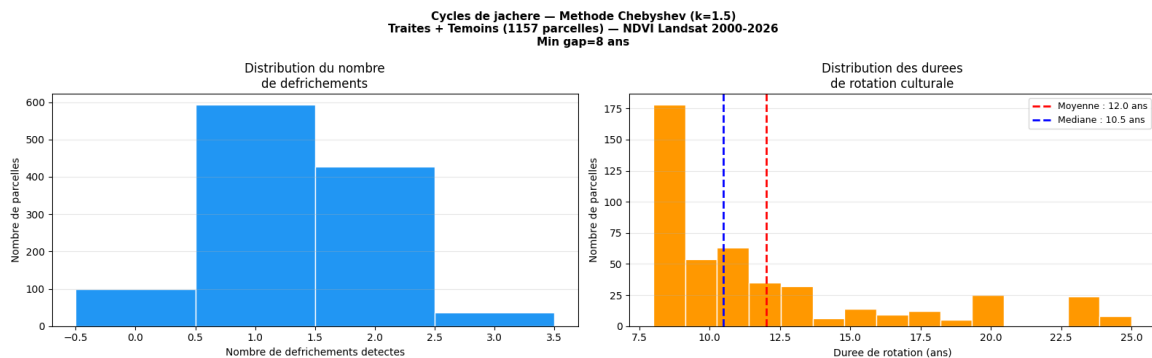
Site	$\sigma$ historique	Chute $\Delta$ NDVI	$k$ limite
Sogboro	0,0236	0,0142	1,45
Borohila	0,0248	0,0165	1,98
Koyola	0,0234	0,0230	1,93

**Table 3.4:** Calibration du multiplicateur  $k$  sur les trois points GPS. Le paramètre  $k$  limite correspond au rapport entre la chute de NDVI pour l'année de défrichement et l'écart-type historique ( $\sigma$ ) depuis l'année 2000

Le seuil  $k = 1.5$  a été retenu. Ce seuil correspond à une probabilité inférieure à 44% que la chute dépasse la variabilité naturelle selon l'inégalité de Tchebychev, et permet de détecter les défrichements détectables à résolution annuelle pour toutes les trois parcelles sans multiplier les fausses alertes.

#### 3.4.2 Durée des cycles de jachère

L'analyse des séries NDVI 2000–2026 par méthode Tchebychev ( $k = 1.5$ ,  $n_{\min} = 8$  ans) détecte une durée moyenne de rotation culturale de 12 ans (médiane 10,5 ans ; Figure 3.3). Sur les 1 157 parcelles, 1 058 (91%) ont connu au moins un défrichement détecté, confirmant une agriculture itinérante active sur l'ensemble du périmètre.



**Figure 3.3:** Distribution de la durée de rotation culturale (intervalle entre deux défrichements successifs, Tchebychev  $k = 1.5$ ,  $n_{\min} = 8$  ans) sur les parcelles Linko et Samana.

La sensibilité du résultat aux paramètres  $k$  et  $n_{\min}$  est illustrée au Tableau 3.5. Le choix  $k = 1.5$  et  $n_{\min} = 8$  ans représente une zone de stabilité relative dans cette matrice et est cohérent avec les cycles documentés en Guinée.

**Table 3.5:** Rotation moyenne (ans) en fonction de  $k$  et  $n_{\min}$ 

$k$	$n_{\min}$ (ans)						
	7	8	9	10	11	12	15
1.0	10,45	11,86	12,68	13,85	14,95	15,82	19,19
1.2	10,56	12,09	13,00	14,29	15,39	16,24	19,42
1.5	10,70	<b>12,02</b>	12,91	14,14	15,16	16,34	19,44
2.0	10,07	11,18	12,49	13,68	14,51	16,31	20,69
2.5	10,67	12,50	12,91	15,14	16,00	21,00	21,00
3.0	7,00	–	–	–	–	–	–

### 3.5 Détection du leakage par Difference-in-Differences

L'analyse DiD a été conduite sur 55 des 62 parcelles de la cohorte 2021 (7 parcelles exclues faute de témoins valides). Les résultats pour les quatre rayons de buffer sont présentés au Tableau 3.6.

Buffer	n	DiD > 0	DiD médian	p-value	Sig.
500 m	55	8 (15%)	−0,00699	< 0,001	***
1000 m	55	9 (16%)	−0,00540	< 0,001	***
2000 m	55	14 (25%)	−0,00225	< 0,001	***
3000 m	55	20 (36%)	−0,00078	0,008	**

**Table 3.6:** Résultats DiD pour quatre rayons de buffer. Test de Wilcoxon bilatéral,  $H_0$  : DiD médian = 0. Le DiD médian est négatif et significatif à tous les buffers.

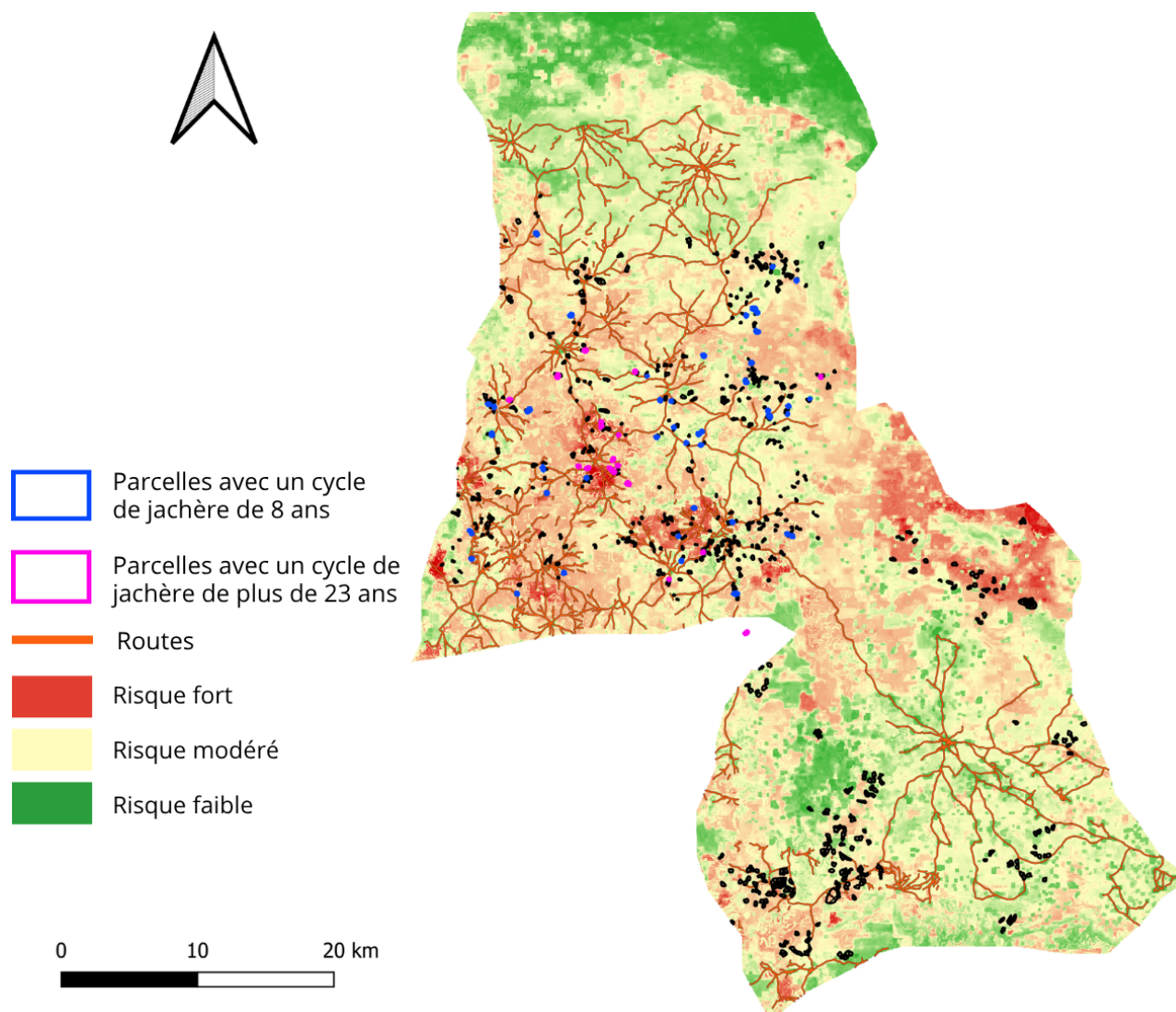
### 3.6 Carte de risque de défrichement

Le modèle Random Forest a été entraîné sur les données historiques de perte de couverture arborée (2001–2024) fournis par JRC-TMF avec sept variables explicatives (Tableau 3.8). Les métriques de performance sont présentées au Tableau 3.7.

Classe	Précision	Rappel	F1-score
Intact	0,996	0,674	0,804
Défrichement détecté	0,020	0,722	0,040

**Table 3.7:** Performance du modèle Random Forest pour la classification des parcelles.

L'importance relative des variables montrent que les caractéristiques du sol (CEC, Argile, pH) dominent avec 62,7% de l'importance totale, suivies par la distance aux routes (19,2%), la pente (8,8%), les zones protégées (5,8%) et la densité de population (3,4%). La Figure 3.4 présente la carte des probabilités prédites par le modèle sur les deux sous-préfectures.



**Figure 3.4:** Carte de risque de défrichement futur (probabilités prédites par Random Forest) pour les sous-préfectures de Linko et Samana.

Variables individuelles		Regroupement par composante	
Variable	Importance (%)	Composante	Importance (%)
CEC	23,60	Caractéristiques du sol <sup>1</sup>	62,72
Argile	20,58	Distance aux routes	19,22
pH du sol	18,54	Pente	8,84
Distance aux routes	19,22	Zones protégées	5,84
Pente	8,84	Densité de population	3,38
Zones protégées	5,84		
Densité de population	3,38		
<b>Total</b>	<b>100,00</b>	<b>Total</b>	<b>100,00</b>

**Table 3.8:** Importance relative des variables explicatives dans le Random Forest.

<sup>1</sup>Somme cumulée des variables pédologiques : CEC, Argile et pH du sol.

## 4 Discussion

### 4.1 Reconstruction des cycles de jachère par méthode Tchebychev

L'analyse par méthode Tchebychev détecte une durée médiane de rotation de 10,5 ans, cohérente avec les cycles d'agriculture itinérante documentés dans cette région. Sur 1,157 parcelles, 91% ont connu au moins un défrichement, confirmant une agriculture itinérante active sur l'ensemble du périmètre.

Cette approche adapte la détection aux variations propres de chaque parcelle, offrant un avantage méthodologique majeur par rapport aux seuils absolus. Cependant, le choix des paramètres  $k = 1.5$  et  $n_{\min} = 8$  est adapté d'une étude dans un contexte malgache [4]. La matrice de sensibilité du Tableau 3.5 montre que les résultats varient significativement selon ces paramètres. Une campagne terrain permettrait de recalibrer robustement les paramètres sur une base étendue (voir Section 4.4).

### 4.2 Détection du leakage par Difference-in-Differences

L'interprétation causale des résultats DiD est limitée par un facteur temporel fondamental : l'horizon d'observation insuffisant. Selon la durée classique de jachère (8-12 ans selon les informations connues pour cette région), les parcelles plantées en 2021 seraient théoriquement disponibles pour un nouveau défrichement à partir de 2029. Les résultats présentés constituent donc une baseline méthodologique validant que la méthode DiD est techniquement applicable, mais non une mesure du leakage effectif.

Le DiD médian négatif et statistiquement significatif à tous les buffers ( $p < 0,01$ ), avec un pourcentage de parcelles présentant un DiD positif variant de 15% à 36%, est cohérent avec cette absence de phase active de leakage : les parcelles ArboRise n'ont pas entraîné un déplacement du défrichement durant 2021–2024.

Une limitation importante mérite attention : 2021 correspondait au lancement du projet, et le caractère nouveau du projet a pu influencer le profil des agriculteurs participants. Les agriculteurs ayant accepté de participer disposaient possiblement de plus de terres (profil socio-économique potentiellement plus aisé), ou la parcelle choisie pour ArboRise n'était pas parmi les plus productives de leur exploitation. Cette auto-sélection pourrait réduire le signal de leakage observable. Une analyse comparative sur les cohortes ultérieures permettrait de tester si ce biais est systématique.

L'application de la méthode DiD à partir de 2029 minimum constituera l'étape critique pour valider empiriquement la présence ou l'absence de leakage.

### 4.3 Carte de risque de défrichement

Le modèle Random Forest identifie les caractéristiques pédologiques (pH, argile, CEC) comme le facteur dominant du risque (62,7%), suivi par la distance aux routes (19,2%), la pente (8,8%), les zones protégées (5,8%) et la densité de population (3,4%). Cette hiérarchie est cohérente, les agriculteurs choisissent rationnellement des terres fertiles pour optimiser les rendements.

Cependant, l'examen des données pédologiques montre que la qualité des sols est globalement homogène et favorable à l'agriculture sur toute la zone. Par conséquent, l'accessibilité et la proximité des routes deviennent le second facteur explicatif majeur. La distance aux routes représente une part significative de l'importance du modèle, confirmant que l'accès aux infrastructures routières est un facteur clé du risque de défrichement.

Les observations régionales confirment cette tendance. À Linko, le risque de défrichement suit les dynamiques d'accessibilité : il est plus élevé près des villages reliés aux axes routiers principaux, qui rayonnent depuis le nord vers l'est et l'ouest. Cette connectivité influence directement les pratiques : les parcelles aux cycles de jachère les plus courts (8 ans) se concentrent le long de ces corridors routiers majeurs. À l'inverse, la partie centrale de Linko, caractérisée par un réseau secondaire de pistes isolées, affiche des cycles plus longs (dépassant 23 ans), reflet d'une pression agricole moindre et d'un temps de régénération des sols plus long.

À Samana, au nord-est de la zone, on observe un risque de défrichement élevé malgré l'isolement routier. Cette région se distingue par une basse capacité d'échange cationique (CEC), particulièrement au centre. Cette anomalie pédologique coïncide avec un espace historiquement dédié à l'élevage plutôt qu'à l'agriculture. Cette affectation des terres explique les modélisations de risque : la probabilité de défrichement y est modérée à faible, car l'élevage ne génère pas les mêmes dynamiques de coupe que l'agriculture. Le risque observé malgré l'isolement suggère que les pratiques agricoles locales et traditions jouent un rôle important, au-delà de la seule accessibilité routière.

L'analyse des performances révèle un comportement typique des modèles appliqués à des données déséquilibrées. Les zones intactes sont prédites avec une précision de 99,6%, très fiable. À l'inverse, la classe « Défrichement détecté » affiche une précision de seulement 2,0%. Cela ne signifie pas que le modèle échoue à détecter les vrais défrichements, son rappel de 72,2% montre qu'il en capture la grande majorité, mais qu'il produit beaucoup de fausses alertes. Dans une perspective d'alerte précoce et de prévention, cette tendance à surestimer le risque est acceptable et même souhaitable : mieux détecter les zones à surveiller que d'en oublier.

L'outil confirme que zones témoins et traitées présentent des profils de risque comparables, soutenant la validité de l'analyse DiD. À court terme, ajouter les centroïdes des villages pour Linko et Samana (données disponibles mais non intégrées faute de temps) améliorerait significativement la capture de l'accessibilité sociale.

#### 4.4 Limitations transversales, solutions et prochaines étapes

Une limite structurelle affecte les trois objectifs : l'absence de données terrain documentant précisément les dynamiques de défrichement à l'échelle familiale. On pourrait théoriquement délimiter les parcelles agricoles en observant les changements de couverture sur les images satellites d'une année à l'autre : un changement de couverture signifierait un défrichement et confirmerait qu'il s'agit d'une parcelle. Cependant, cette approche est très chronophage et ne produit pas une délimitation précise. Une campagne terrain documentant directement les parcelles et leurs cycles paraît donc plus adaptée et efficace. Pour la méthode Chebyshev, les trois points GPS actuels sont insuffisants pour calibrer robustement les paramètres régionaux. Pour le DiD, l'absence de documentation du stade de cycle des parcelles témoins introduit une hétérogénéité incontrôlée.

Fondamentalement, la télédétection seule ne peut pas mesurer le leakage sans données documentant la localisation exacte des terres familiales cultivées. Les buffers spatiaux constituent un proxy utile mais imparfait. De plus, Hansen mesure la perte de couverture arborée sans distinguer défrichements agricoles et déforestation forestière, signifiant que tous les résultats incluent potentiellement un bruit provenant de dynamiques forestières non liées au projet ArboRise.

Une autre limitation importante concerne les données utilisées pour identifier les défrichements. Le produit Tropical Moist Forests (TMF) du JRC a été calibré à l'échelle mondiale sur plus de 70 pays. Bien que TMF soit capable de détecter la dégradation forestière, son application en Guinée est peu adaptée. Il a été conçu pour suivre les forêts naturelles, pas les zones agricoles complexes. De plus, TMF divise la couverture végétale en plus de 40 classes différentes. Cela semble riche en théorie, mais choisir et regrouper ces classes dans un paysage agricole est difficile sans une bonne connaissance du terrain.

Pour pallier ce manque de spécificité agricole, une tentative de comparaison croisée a été menée avec le produit Copernicus Global Land Cover afin d'isoler rigoureusement les espaces cultivés. Toutefois, l'examen des secteurs de Linko et de Samana, pourtant empiriquement reconnus comme majoritairement anthropisés et dédiés à l'agriculture, a révélé une anomalie de classification, le modèle de Copernicus répertoriant ces espaces comme des forêts. Cette incohérence illustre l'insuffisance des produits disponibles pour différencier défrichement agricole et déforestation forestière.

L'analyse révèle également que certaines zones géographiquement isolées affichent un risque de défrichement accru sans corrélation évidente avec l'accessibilité routière, suggérant que les modèles basés sur l'accessibilité physique omettent les facteurs socio-économiques et culturels : pratiques agricoles

locales, traditions d'exploitation, dynamiques communautaires ne sont pas capturées par les variables géospatiales.

Une seule campagne terrain peut simultanément satisfaire les besoins des trois objectifs et permettre une validation empirique de Hansen. Pour chaque parcelle témoin sélectionnée et représentant les deux zones et différents contextes d'accessibilité, les informations suivantes seraient pertinentes :

1. Géoréférencer précisément le polygone de la parcelle
2. Documenter les dates précises des derniers défrichements et les durées de jachère qui ont suivi, permettant de recalibrer les paramètres  $k$  et  $n_{\min}$  de Chebyshev et de calculer le stade actuel du cycle
3. Identifier et géoréférencer si possible les autres parcelles cultivées par la famille, permettant de vérifier où peut survenir le leakage.
4. Pour un échantillon de pixels identifiés comme défrichés par Hansen, vérifier sur le terrain s'il y a réellement du défrichement (détection des faux positifs et faux négatifs)

Plus le nombre de parcelles témoins est élevé, plus robustes seront la calibration de Chebyshev et la validation de Hansen.

À plus long terme, l'accès à des produits géospatiales de meilleure spécificité agricole affineront la robustesse des analyses. La nouvelle version de Hansen prévue pour 2026, capable de détecter les pixels ayant connu plusieurs pertes de couverture arborée à des années différentes, s'avérera particulièrement pertinente pour cette région : elle permettra une caractérisation fine des cycles de jachère et une meilleure discrimination entre agriculture itinérante et déforestation. La réapplication de la méthode DiD à partir de 2029 permettra de valider empiriquement la présence ou l'absence de leakage, transformant le cadre méthodologique actuel en résultats avec implications causales directes pour l'évaluation du projet ArboRise.

## 5 Conclusion

Ce rapport présente une approche méthodologique pour évaluer le leakage de carbone associé au projet ArboRise en Guinée. La combinaison de trois analyses (reconstruction des cycles de jachère, analyse Difference-in-Differences, et cartographie du risque) offre un cadre rigoureux pour mesurer empiriquement la présence ou l'absence de leakage. Cependant, cette étude soulève d'abord une limite structurelle : la télédétection seule ne suffit pas. Sans données terrain documentant précisément les parcelles familiales et leurs historiques de défrichement, une attribution causale du leakage n'est pas possible.

L'horizon critique pour mesurer un leakage réel est 2029. Avant cette date, toute analyse constitue une baseline méthodologique validant la faisabilité technique, non une mesure d'impact. Les agriculteurs participants au projet pourraient également disposer de ressources supérieures aux non-participants, réduisant le signal observable. Ces considérations temporelles et structurelles expliquent pourquoi les résultats actuels, bien que statistiquement significatifs, ne permettent pas de conclure sur un leakage effectif.

Les trois objectifs révèlent aussi comment les facteurs géographiques et socio-économiques interagissent. L'accessibilité routière et la qualité pédologique des sols expliquent une part majeure du risque de défrichement, mais les zones isolées de Samana montrent un risque élevé malgré l'absence d'infrastructure. Cela suggère que les modèles géospatiales capturent mal les pratiques agricoles locales, les traditions d'exploitation, et les dynamiques communautaires, des variables impossibles à mesurer par satellite.

Une campagne terrain serait l'étape suivante idéale. Elle permettrait simultanément de calibrer la méthode de détection des cycles sur une base étendue de parcelles documentées, stratifier l'analyse DiD par étape de cycle comparable entre groupes, valider Hansen en quantifiant ses erreurs, et affiner la carte de risque. Par ailleurs, la nouvelle version de Hansen prévue pour 2026 permettra une meilleure

discrimination entre agriculture itinérante et déforestation. La réapplication de la méthode DiD à partir de 2029 permettra enfin de valider empiriquement le leakage.

Au-delà des résultats techniques, cette étude illustre une tension plus profonde. La tendance régionale à la hausse de la perte de couverture arborée, observable depuis 2015 et indépendante du projet, rappelle que les moteurs structurels de la déforestation (pression démographique, appauvrissement des sols, rareté des terres fertiles) ne sont pas adressés par la reforestation parcellaire seule. En l'absence de conditions permettant aux familles de réduire durablement leur dépendance à l'agriculture itinérante, le leakage constitue une réponse rationnelle des ménages face à une contrainte imposée sur leurs terres productives. La rigueur méthodologique dans l'évaluation de ces projets reste un prérequis indispensable à tout jugement éclairé sur leur contribution réelle aux objectifs climatiques.

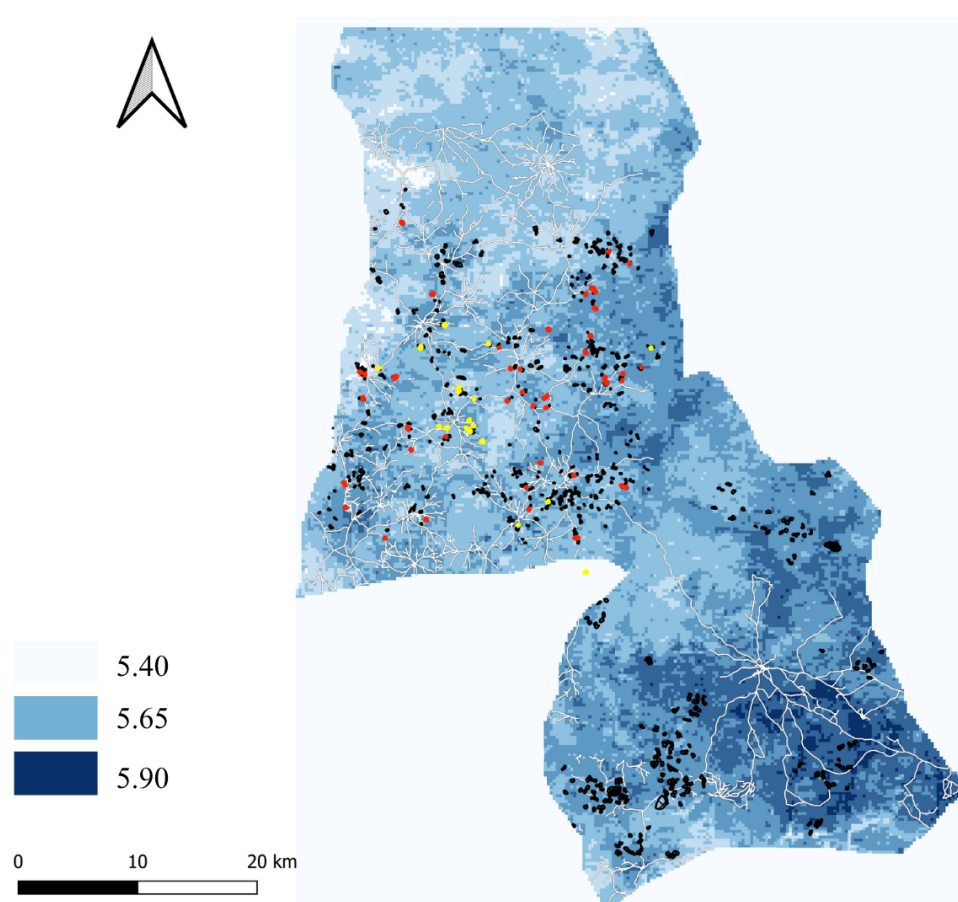
## Annexe

## Comparaison Hansen / NDVI sur plusieurs années

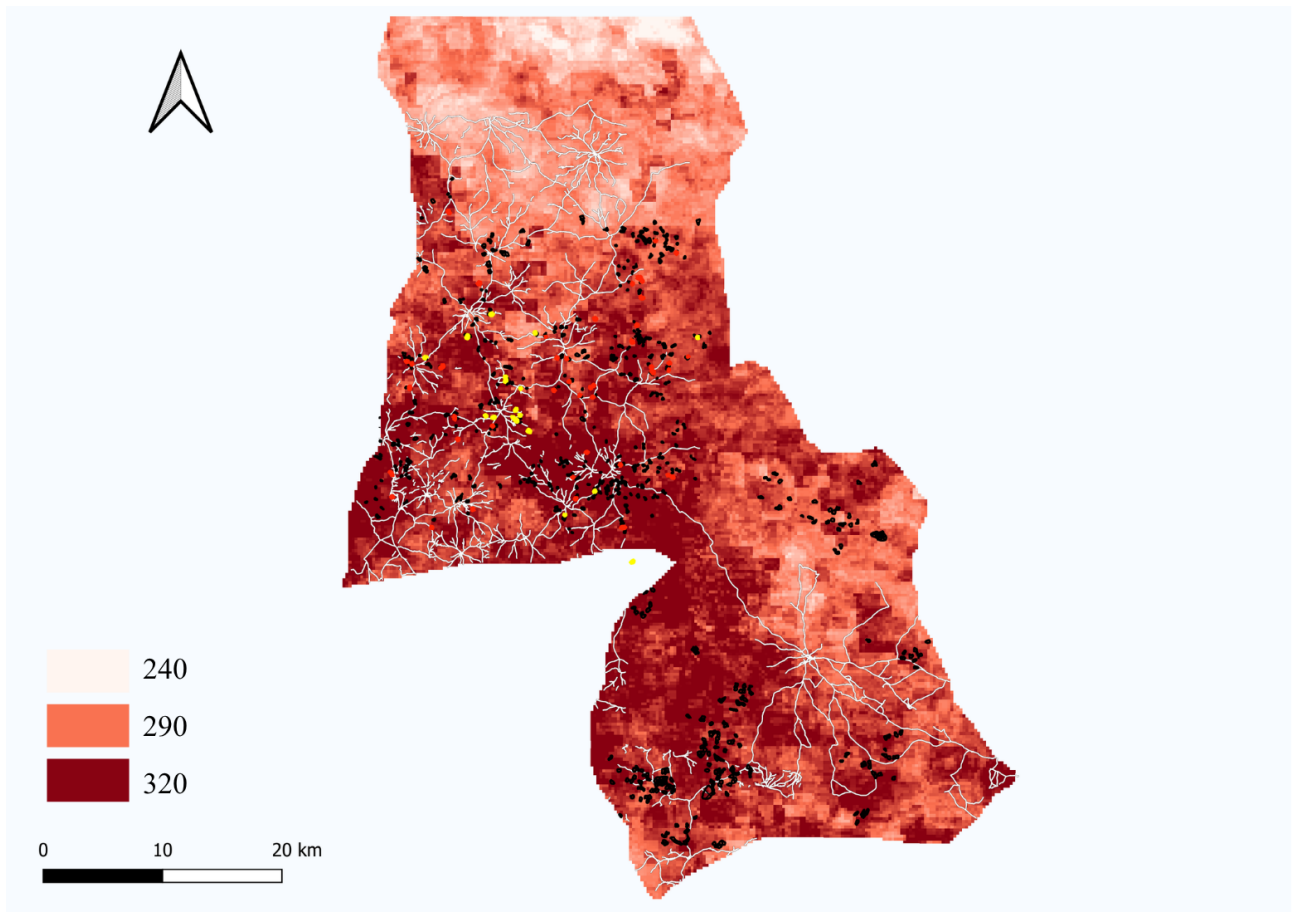
Année	Hansen détecte	NDVI détecte	S1	S2	S3	S4	Accord
2017	67	4	4	67	1086	0	0,0%
2018	57	60	50	47	1050	10	17,5%
2019	34	15	11	30	1112	4	11,8%
2020	51	26	17	42	1089	9	17,6%
2021	13	1	0	12	1144	1	7,7%

**Table 5.1:** Comparaison Hansen vs NDVI-dérivée (seuil 0.034) pour les années 2017–2021 sur l'ensemble des 1 157 parcelles

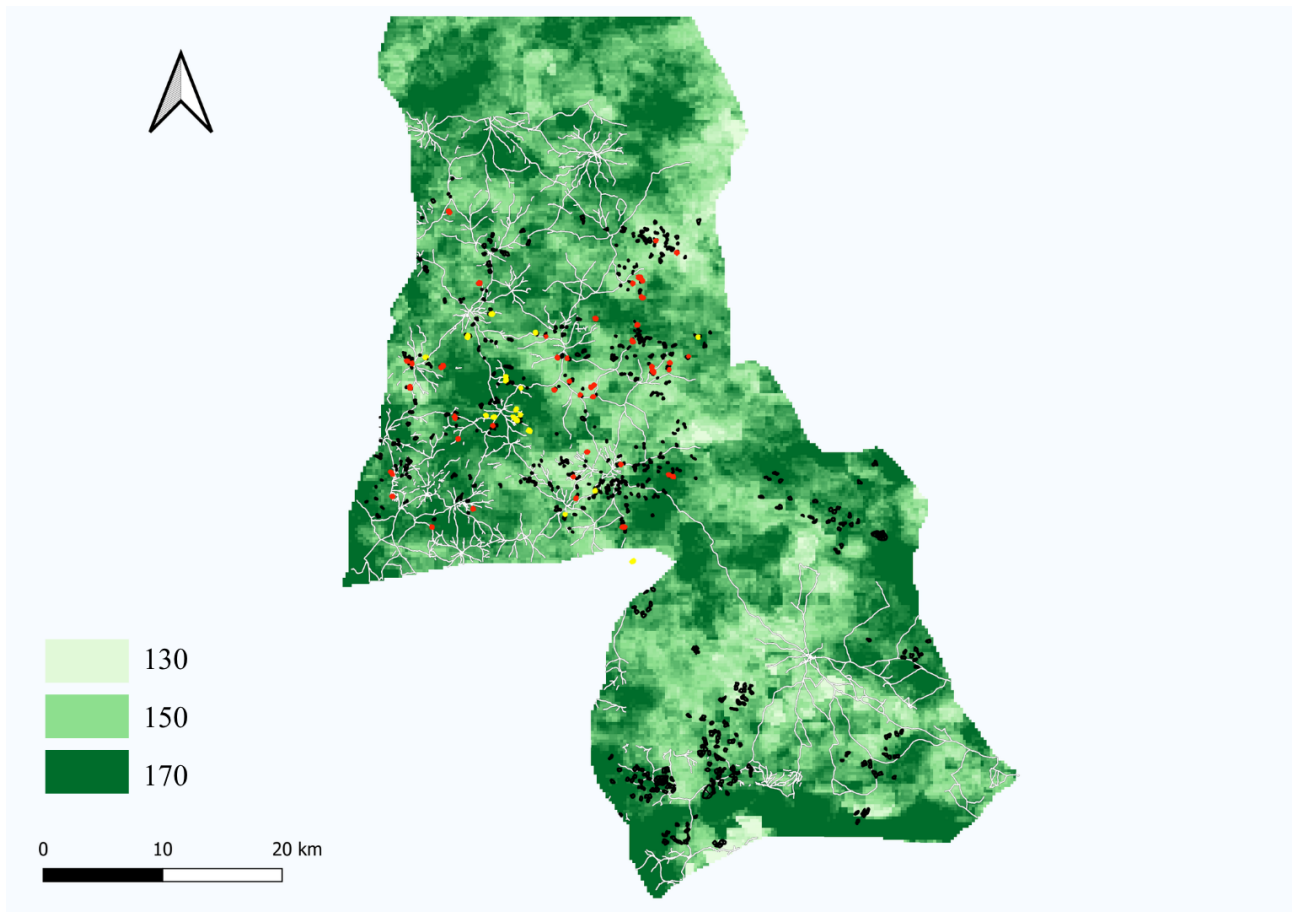
## Variables pédologiques et cycles de jachère



**Figure 5.1:** Carte du pH des sols de Linko et Samana avec les durées de cycle de jachère (jaune : Cycle > 23 ans , rouge : cycle de 8 ans )



**Figure 5.2:** Carte de la teneur en argile des sols de Linko et Samana avec les durées de cycle de jachère (jaune : Cycle > 23 ans , rouge : cycle de 8 ans )



**Figure 5.3:** Carte de la capacité d'échange cationique (CEC) des sols à Linko et Samana avec les durées de cycle de jachère (jaune : Cycle > 23 ans , rouge : cycle de 8 ans )

## References

- [1] Daniel S. Silva and Samia Nunes. “Leakage Effects from Reforestation: Estimating the Impact of Agricultural Displacement for Carbon Markets”. In: *Land* 14.5 (2025), p. 963. DOI: 10.3390/land14050963.
- [2] Google Earth Engine Team. *Google Earth Engine: A Planetary-Scale Platform for Earth Science Data and Analysis*. 2024. URL: <https://earthengine.google.com>.
- [3] M. C. Hansen et al. “High-Resolution Global Maps of 21st-Century Forest Cover Change”. In: *Science* 342.6160 (2013), pp. 850–853. DOI: 10.1126/science.1244693.
- [4] Federica Ghilardi et al. “A Possible Role of NDVI Time Series from Landsat Mission to Characterize Lemurs Habitats Degradation in Madagascar”. In: *Science of the Total Environment* 974 (2025), p. 179243. DOI: 10.1016/j.scitotenv.2025.179243.