

# THÈSE POUR OBTENIR LE GRADE DE DOCTEUR DE L'UNIVERSITÉ DE MONTPELLIER

En Géomatique

École doctorale GAIA – Biodiversité, Agriculture, Alimentation, Environnement, Terre, Eau – n°584  
Portée par l'Université de Montpellier

Unité de recherche : UMR Espace-DEV/ UMR TETIS

## From Land Cover to Land Use Systems Mapping: Detection and Characterization of Large Scale Agricultural Investments (LSAIs) from Satellite Imagery. Application to Senegal

Présentée par Yasmine NGADI SCARPETTA  
Le 9 avril 2024

Sous la direction de :  
Anne-Elisabeth LAQUES et Agnès BÉGUÉ

Devant le jury composé de

Damien ARVOR, Chargé de Recherche, CNRS

Nicolas DELBART, Professeur des Universités, Université Paris Cité

Liam WREN-LEWIS, Chargé de Recherche, INRAE

Carmen GERVET, Professeure des Universités, Université de Montpellier

Agnès BÉGUÉ, Cadre Scientifique des EPIC, CIRAD

Valentine LEBOURGEOIS, Cadre Scientifique des EPIC, CIRAD

Invitée :

Anne-Elisabeth LAQUES, Directrice de Recherche, IRD

Rapporteur

Rapporteur

Examineur

Présidente du jury

Co-Directrice de thèse

Encadrante

Directrice de thèse



UNIVERSITÉ  
DE MONTPELLIER

# ABSTRACT

---

Increasing demand for water, food and energy has led to dramatic competition for land, resulting in a global land rush in the form of large scale agricultural investments (LSAI). Due to their many potential negative impacts and the opacity of their surroundings, accurate detection and characterisation of LSAIs in space and time is required. The increasing availability of dense satellite imagery time series (SITS), together with ever-improving change detection algorithms, is useful in this task. While SITS change detection algorithms are efficient at detecting abrupt and gradual changes phenological time series, there is still much room for improvement when it comes to detecting seasonal changes.

The primary objective of this research was to automatically detect, in an unsupervised manner, the implementation of LSAIs in Senegal based on remote sensing data. This work is structured around three interrelated papers. The first presents a fast and unsupervised approach (BFASTm-L2) developed to detect, in full MODIS NDVI SITS at the pixel level, the breakpoint associated with the largest *pattern* (i.e. mostly seasonal) change of the time series. Compared to other change detection algorithms (BFAST Lite, EDYN and BFAST monitor), BFASTm-L2 proved to be particularly sensitive to seasonal changes and efficient in highlighting LSAIs in Senegal. This supports the hypothesis that changes induced by land use systems such as LSAIs are very often of a seasonal type. The second paper sought to differentiate the contribution of LSAIs from the main drivers of change (climatic, natural and anthropogenic) at a national-scale, relying mainly on three time series-based change metrics calculated at the pixel level (magnitude of change, direction of change, dissimilarity), which, when combined into a unique composite map, provided insights into land dynamics. LSAIs were shown to have a specific ecoregional signature of change. Finally, the third paper aims to refine the detection of the deals by automatically locating potential hotspots of change related to LSAIs in two contrasting ecoregions of Senegal through the segmentation of a BFASTm-L2-based magnitude of change map combined with object-based K-means clustering. In this last study, key discriminative metrics (textural and structural) derived from higher resolution imagery (Landsat) were combined with the spectro-temporal ones coming from MODIS NDVI SITS to provide a generic characterization of LSAIs.

Through its specific focus on large-scale detection of LSAIs, this project contributed to the land change community by improving the understanding of land dynamics and the drivers of change behind the detected changes.

**Keywords: SITS, MODIS NDVI, land use and land cover change, BFASTm-L2, LSLA, unsupervised change detection, change metric**

## RÉSUMÉ

---

Pendant les deux dernières décennies, la demande croissante en eau, en alimentation et en énergie a généré une ruée mondiale vers les terres sous la forme d'investissements agricoles à grande échelle (IAGE). En raison de leurs nombreuses incidences potentiellement négatives et de l'opacité entourant les données disponibles, la détection et la caractérisation de ces IAGEs dans le temps et dans l'espace sont nécessaires. La disponibilité croissante de séries temporelles d'images satellitaires (STIS), associée à l'amélioration constante des algorithmes de détection des changements, est particulièrement intéressante dans cette tâche. Bien que les algorithmes de détection des changements soient efficaces dans la détection des changements brusques et graduels dans les séries temporelles phénologiques, des progrès restent nécessaires en ce qui concerne la détection des changements saisonniers.

L'objectif principal de cette recherche est la détection automatique, de manière non supervisée, des IAGEs implantés au Sénégal à partir des données de télédétection. Ce travail est structuré autour de trois articles interdépendants. Le premier présente une méthode rapide et non supervisée (BFASTm-L2) développée pour détecter, dans les STIS de NDVI issues du satellite MODIS, le point de rupture associé au plus grand changement de forme (principalement saisonnier) dans la série temporelle de chaque pixel. Comparé à d'autres algorithmes de détection des changements (BFAST Lite, EDYN et BFAST monitor), BFASTm-L2 s'est révélé particulièrement sensible aux changements saisonniers et efficace pour mettre en évidence les IAGEs au Sénégal. Cela confirme l'hypothèse selon laquelle les changements d'utilisation des terres tels que ceux générés par l'implantation de IAGEs sont très souvent de type saisonnier. Le deuxième article a cherché à différencier la contribution des IAGEs de celles des principaux facteurs de changement (climatiques, naturels et anthropiques) dans les changements détectés. L'approche mise en place s'est basée sur l'identification de trois variables spectro-temporelles issues des séries temporelles de NDVI MODIS de chaque pixel (la magnitude du changement, la dissimilarité induite par le changement et la direction du changement), lesquelles, une fois combinées dans une image composite RVB, ont permis l'obtention d'un meilleur aperçu de la dynamique des terres et de leur possible facteur de changement. Il a ainsi été démontré que les IAGEs ont une signature de changement particulière au niveau éco-régional. Enfin, le troisième article vise principalement affiner la détection des transactions en localisant automatiquement les hotspots de changement potentiellement liés aux IAGEs dans deux écorégions contrastées du Sénégal. Cette

localisation s'opère grâce à la segmentation d'une carte des magnitudes de changement basée sur BFASTm-L2, combinée à un clustering K-means à l'échelle de l'objet. Dans cette dernière étude, des métriques discriminantes clés texturales et structurelles), dérivées d'images à plus haute résolution (Landsat), ont été combinées aux métriques spectro-temporelles issues des séries temporelles de NDVI MODIS pour l'étape de clustering, et ainsi fournir une caractérisation des IAGEs.

En privilégiant la détection non supervisée et l'analyse du type de changement induit dans les séries temporelles phénologiques, cette recherche contribue significativement à une meilleure compréhension de la dynamique des terres et des facteurs de changement à l'origine des modifications détectées.

**Mots clés: Séries Temporelles d'Images Satellitaires, SITS, MODIS NDVI, changement d'utilisation des sols, IAGE, BFASTm-L2, détection non supervisée de changements, métriques de changement**

**Titre :** De l'image satellite au système d'utilisation des terres: détection et caractérisation des investissements agricoles à grande échelle (IAGE) à partir de données d'Observation de la Terre. Etude de cas du Sénégal



# ACKNOWLEDGEMENTS

---

First of all, I am deeply grateful to my supervisors and co-supervisors Agnès Bégué, Anne-Elisabeth Laques and Valentine Lebourgeois for giving me the opportunity to work in this field of land research, which is of particular importance to me. Their constant availability, support, open-mindedness, patience, cheerfulness and most importantly, their trust in me, together with the freedom and flexibility given to me in my work, have been indispensable for the successful completion of this thesis.

My thanks extend to the members of the PhD Jury, the members of my Individual Thesis Comitees the partners of ISRA-BAME in Senegal for generously giving their time and expertise.

Finally, I would like to thank all the people who have indirectly contributed to this thesis. A special mention goes to my family for their great patience, constant support and love, without which this simply would not have been possible.

« Land can be a major engine of shared prosperity  
or one of the most pervasive drivers of inequality”

*Guereña and Wegerif, 2019*

# CONTENTS

---

Abstract .....	2
Résumé .....	3
Acknowledgements .....	5
Contents .....	7
Table of figures.....	9
Table of tables.....	11
Acronyms.....	12
Résumé étendu en français .....	13
Introduction .....	13
Détection non supervisée de changements d'utilisation des terres basés sur les changements saisonniers : BFASTm-L2 .....	18
Cartographie des types de changements par analyse des points de rupture dans les séries temporelles d'images.....	20
Cartographie et caractérisation des investissements agricoles à grande échelle au Sénégal .....	22
Discussion et conclusions .....	24
1 Introduction .....	27
1.1 What are large-scale agricultural investments (LSAIs)? .....	27
1.2 Review of existing satellite-based methods to detect LSAIs and related features....	32
1.3 Identified challenges in remote sensing to detect LSAI related changes .....	38
1.4 Research objective .....	44
2 BFASTm-L2, an unsupervised LULCC detection based on seasonal change detection – An application to Large-Scale Land Acquisitions in Senegal.....	50
2.1 Highlights .....	51
2.2 Abstract .....	51
2.3 Introduction .....	52
2.4 Material and Methods .....	54
2.5 Results .....	67

2.6	Discussion.....	79
2.7	Conclusions and perspectives.....	83
3	Insight into large-scale LULC changes and their drivers through breakpoint characterization – An application to Senegal.....	84
3.1	Highlights.....	85
3.1	Abstract.....	85
3.2	Introduction.....	86
3.3	Data and methods.....	88
3.4	Results and Discussion.....	94
3.5	Discussion.....	107
3.6	Conclusions and perspectives.....	110
3.2	Appendices.....	111
4	Mapping and characterization of Large-Scale Agricultural Investments in Senegal.....	118
4.1	Highlights.....	119
4.2	Abstract.....	119
4.3	Introduction.....	120
4.4	Data and Methods.....	123
4.5	Results.....	133
4.6	Discussion.....	155
4.7	Conclusions.....	161
4.8	Appendices.....	163
5	Discussion.....	170
5.1	Key findings.....	171
5.2	Synthesis, recommendations and perspectives.....	179
	References.....	182
	Appendices.....	192
	Appendix A : A sample of LSAI-related study cases.....	192
	Appendix B : Case Studies from Laos and Argentina.....	197

# TABLE OF FIGURES

---

Figure 1.1: Cumulative global contract size under production (left axis, bars) and share of concluded deals size under production (right axis, dashed line)(adapted from Lay et al. (2021))	27
Figure 1.2: Global heat map of agricultural LSLAs contained in the Land Matrix in 2016. ....	28
Figure 1.3: Status of reported LSLAs in Senegal from different sources.....	31
Figure 1.4: Distribution of on-going and intended agricultural LSLAs in the three main hotspots in Senegal.....	46
Figure 1.5: Thesis conceptual framework.....	47
Figure 2.1: Flowchart of the data and methods.....	56
Figure 2.2. Study area (red box) and study cases (red dots) in Senegal .....	58
Figure 2.3. Flowchart of the BFASTm-L2 approach.....	65
Figure 2.4. Temporal distributions of the highest-absolute magnitude breakpoint detected by BFASTmonitor.....	69
Figure 2.5. Violin plots of the normalized highest-magnitude breakpoint, per type of change .....	70
Figure 2.6. Application of the change detection algorithms to four Senegalese study cases	73
Figure 2.7. Change detection maps for each of the four change detection algorithms. ....	77
Figure 2.8. Running time mean and standard-deviations (5 runs) over areas with varying-size .....	79
Figure 3.1: Flowchart for identifying drivers of change from MODIS time series. ....	88
Figure 3.2: Senegal's map of the MODIS NDVI 2000-2021 average. ....	89
Figure 3.3: Illustration of the highest-intensities changes found in the simulated dataset, for each change-type group .....	91
Figure 3.4: Violin plots of the NDVI ratio for each type of change in the simulated dataset. .	95
Figure 3.5: Selected pixel-study cases .....	96
Figure 3.6: National maps of the different change metrics.....	100
Figure 3.7: RGB composite map.....	104
Figure 3.8: Zoom-ins of study cases 1 to 13. ....	105
Figure 4.1: Three-step research workflow .....	124
Figure 4.2: Overview of the study areas.....	126
Figure 4.3: Boxplots of LSAI size (in hectares) by study region .....	127
Figure 4.4: Schematic view (to scale) of the 100 x 100 pixel Landsat image (full square). Coloured areas were used to identify the composite image with the highest contrast,	

representing the largest difference between the average NDMI value in the red area and the average NDMI value in the blue area. ....	131
Figure 4.5: Change maps for the Niayes (a) and the Senegal River regions (b).....	134
Figure 4.6 Histogram of the relative differences (in %) .....	136
Figure 4.7: Study regions with LSAs from the non-exhaustive ISRA field database (in cyan) and extracted hotspots of change (yellow). ....	138
Figure 4.8: Close-up views of some of the identified change objects.....	140
Figure 4.9:Box plots of the size of detected objects per land use type and study region versus their area. ....	141
Figure 4.10: Pairwise scatterplots and univariate distributions (diagonal) of the spectro-temporal variables .....	142
Figure 4.11:NDMI monthly means for February based on 2019-2021 LANDSAT data .....	143
Figure 4.12: NDMI Landsat-based 3-year composites (100 x 100 pixels).....	145
Figure 4.13: Segmentations (after image thresholding) obtained using the NDMI composites .....	148
Figure 4.14: Univariate distributions of three structural features .....	149
Figure 4.15:K-means clusters (applied to the extracted hotspots) obtained in the Niayes region .....	150
Figure 4.16: Barplots of the K-means cluster composition for the Niayes .....	151
Figure 4.17: Boxplot of clusters (x-axis) in the Niayes against original spectral-temporal variables .....	151
Figure 4.18: K-means clusters (applied to the extracted hotspots) obtained in the SR region .....	153
Figure 4.19: Barplots of the K-means clusters composition for the SR .....	153
Figure 4.20: Boxplot of clusters (x-axis) in the SR against original spectro-temporal variables .....	154
Figure 4.21: Tracking the destination of LSAs reported in the field database.....	155

# TABLE OF TABLES

---

Table 2.1.Types of change, levels of intensities and number of samples present in the simulated data set .....	57
Table 2.2 : Main characteristics of the change-detection algorithms used in this study, their strengths and limitations .....	62
Table 2.3: Average breakpoint magnitude (mean $\pm$ standard deviation) in LSLAs .....	78
Table 3.1: Medians of the dissimilarities computed per type of change using the simulated dataset.....	94
Table 3.2: Sensitivity of the change metrics to different types of change.....	98
Table 3.3: Association table between the RGB map colours (first column) and the potential drivers of change.....	106

# ACRONYMS

---

BFAST	Breaks For Additive Season and Trend
BFASTm	BFAST monitor
EWMACD	Exponentially Weighted Moving Average Change Detection
EO	Earth Observation
GEOBIA	Geographical Object Based Image Analysis
L2	L2 norm Euclidean distance
LOS	Length Of Season
LSLA	Large Scale Land Acquisition
LSAI	Large Scale Agricultural Investment
LULC	Land Use and Land Cover
MODIS	MODerate resolution Imaging Spectroradiometer
NDVI	Normalized Difference Vegetation Index
NGO	Non-Governmental Organization
NOS	Number Of Seasons
SITS	Satellite Imagery Time Series
VI	Vegetation Index



# RÉSUMÉ ÉTENDU EN FRANÇAIS

---

## INTRODUCTION

Depuis la fin des années 2000, les acquisitions de terre à grande échelle (ATGE ou *Large Scale Land Acquisition* en anglais) ont connu une forte augmentation alimentée par la volatilité croissante des marchés des matières premières, une population mondiale en croissance rapide, des évolutions dans les modes de consommation et le changement climatique. Ces acquisitions foncières à grande échelle, souvent appelées "accaparements de terres", font référence à l'occupation de vastes étendues de terres par des individus, des États ou des sociétés (nationales ou étrangères), principalement à des fins agricoles, mais également pour la sylviculture, le tourisme, la conservation, l'exploitation minière, l'expansion urbaine ou de grands travaux d'infrastructure. Les données rapportées montrent une forte concentration géographique des investissements, la majorité des acquisitions de terres arables étant concentrées dans le Sud global et particulièrement en Afrique subsaharienne.

On estime que les transactions ATGE (ou IAGE pour Investissements Agricoles à Grande Echelle) pour la production végétale représentent 83 % de toutes les transactions dans le monde, soit l'équivalent d'environ 27 millions d'hectares en 2016 (Nolte et al., 2016) . Différentes études montrent que la plupart des projets d'IAGE ont des investisseurs étrangers (principalement de l'UE et d'Asie) et ciblent la production de biocarburants (63 %), en particulier le Jatropha. Cependant, n'atteignant souvent pas la rentabilité escomptée, de nombreux projets ont fait faillite ou suspendu leurs opérations, réduit leur surface ou changé de culture, démontrant l'énorme dynamique spatio-temporelle de ces acquisitions. Toujours d'après la Land Matrix (2021), entre 30% et 73 % des terres sous contrat ont été mises en production. Cette incertitude sur les chiffres montre la grande méconnaissance sur le nombre, l'état et la répartition géographique des IAGEs dans le monde.

Les investissements agricoles à grande échelle peuvent affecter négativement la sécurité alimentaire, bouleverser les moyens de subsistance ruraux, notamment en affectant des zones coutumières ou protégées souvent sans possibilité de retour à leur état initial, et dégrader l'environnement par des pratiques agricoles intensives (monocultures, produits agrochimiques, irrigation, mécanisation...) (D'Odorico et al., 2017; Schoneveld, 2014). Parallèlement à ces impacts négatifs, certains auteurs ont souligné les faibles opportunités d'emploi que ces investissements peuvent générer en raison de leur fort taux de mécanisation (Davis et al., 2014). L'importance croissante des IAGEs a suscité de nombreuses études pour évaluer leurs impacts socio-économiques et environnementaux, mais la quantification précise de ces investissements reste difficile en raison de données peu fiables et difficiles d'accès. Il

est donc urgent de développer des systèmes capables de détecter, caractériser et surveiller automatiquement les acquisitions de terres à grande échelle dans l'espace et le temps. A cette fin, les données d'observation de la terre représentent un outil intéressant pour la détection et la caractérisation des IAGE.

## Etat de l'art des méthodes en télédétection pour détecter les IAGE

Alors que la détection des changements d'occupation des sols a été largement étudiée, des progrès restent à faire pour détecter les changements plus complexes d'utilisation des terres (Verburg et al., 2019). Jusqu'à présent, la plupart des recherches se sont concentrées sur le développement d'approches supervisées pour détecter les IAGEs ou les caractéristiques associées aux IAGEs basées sur des techniques d'apprentissage automatique, et plus récemment, avec la disponibilité croissante de données à très haute résolution spatiale, sur des techniques d'apprentissage profond. Les approches proposées pour détecter les systèmes agricoles intensifs sont principalement basées sur : (1) des schémas phénologiques distincts (souvent basés sur le nombre de cycles de culture), (2) les spécificités des objets du paysage en termes de taille et d'arrangement dans l'espace (taille des champs, géométrie des champs, homogénéité...). Ces techniques reposent cependant sur la disponibilité d'un jeu de données d'entraînement vaste, fiable et représentatif. Or, les bases de données existantes sur les IAGEs sont incomplètes et les données de référence sont souvent obtenues par interprétation visuelle d'images satellites. Les approches supervisées sont aussi particulièrement sensibles aux conditions d'acquisition des images, et sont souvent liées à une région et/ou à une culture spécifique, avec des hypothèses difficilement transférables à d'autres zones d'étude. Enfin, les études existantes se concentrent généralement sur des régions qui subissent un seul type de changement (déforestation par exemple).

Ces nombreuses limites justifient le besoin d'approches non supervisées, basées sur la détection de caractéristiques spatio-temporelles et spectrales induites par les changements dans les systèmes d'utilisation des terres. Bien que cruciaux pour détecter les schémas de changement, ces algorithmes peinent à repérer les nuances des changements d'utilisation des terres en raison des variations saisonnières, liées au climat ou à d'autres facteurs, masquant souvent des changements saisonniers plus significatifs. Ainsi, distinguer les variations dues aux IAGEs des autres facteurs de variation naturels ou anthropiques représente un défi crucial pour une gestion des ressources efficace et des prises de décision éclairées.

En résumé, il y a un besoin croissant de méthodes non supervisées et génériques, pour s'adapter à la diversité des systèmes agricoles et au suivi de leurs changements. L'analyse des séries temporelles d'images satellites longues et denses (SITS) offre de grandes

perspectives à cet égard, car elles peuvent fournir de nouveaux éclairages sur la dynamique des systèmes agricoles intensifs.

## Objectifs et objets de recherche

L'objectif principal de cette thèse est d'**explorer le potentiel des séries temporelles d'images de télédétection pour détecter et caractériser de manière non supervisée les investissements agricoles à grande échelle (IAGEs) à différentes échelles.**

Les IAGEs sont ici définis comme des exploitations agricoles compactes, avec une surface minimale de 30 hectares et pratiquant l'agriculture intensive (mécanisée, avec de forts intrants), quel que soit leur régime foncier (par exemple, les agro-industries, les coopératives, etc.). La portée de cette étude, tout en cherchant à être aussi générale que possible, est limitée au Sénégal en tant que démonstration de faisabilité. Le choix de ce pays est lié à la disponibilité d'une base de données de référence sur les ATGEs (incluant les IAGEs), et à la diversité de ses éco-régions. Trois éco-régions contrastées, et concentrant 92% des transactions foncières en cours et 100% des installations actuelles et prévues dans le pays (Bourgoin et al., 2019), ont été particulièrement étudiées : la vallée du fleuve Sénégal dans le nord, la zone sylvo-pastorale du Ferlo et une zone étendue des Niayes, près de Dakar (Figure 1). La période d'étude, couvre la disponibilité temporelles des données MODIS et est comprise entre 2000 et 2021.

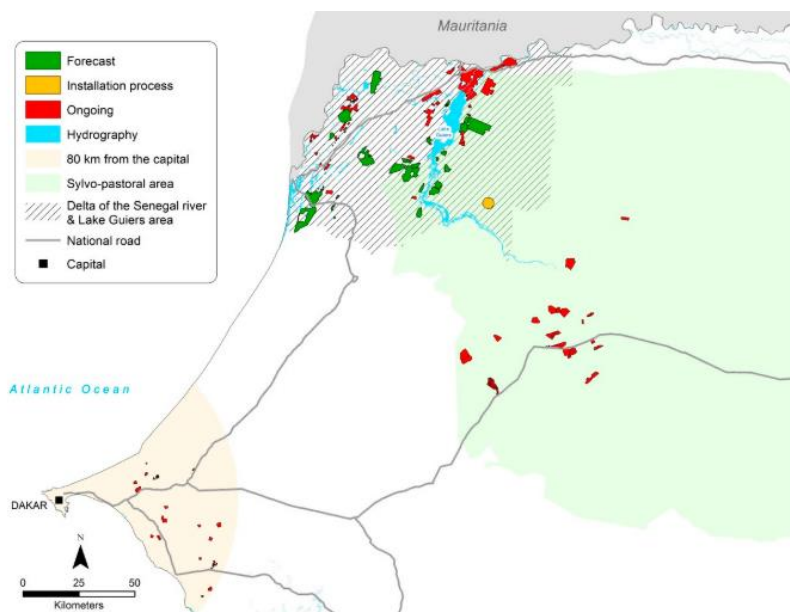


Figure 0. Répartition des investissements agricoles à grande échelle (IAGEs) en cours et prévus dans les trois principaux hotspots au Sénégal : la vallée du fleuve Sénégal (zone hachurée), le Ferlo (zone verte) et la zone étendue des Niayes (zone orange). Source: (Bourgoin et al., 2019).

## Cadre conceptuel et structuration de la thèse

Etant attendu que les IAGEs mis en production (i) induisent des changements d'utilisation des terres qui peuvent se refléter dans les séries temporelles d'indices de végétation, (ii) occupent de vastes zones et (iii) pratiquent une agriculture intensive, nous avons développé un cadre conceptuel de détection des IAGEs à partir de variables de télédétection basées sur ces caractéristiques. La stratégie adoptée découlant de ce cadre conceptuel se compose de trois étapes principales (Figure 2) :

(1) La première étape consiste en la détection automatique et la sélection, au sein de longues et denses séries temporelles d'indices de végétation, du changement qui est le plus susceptible d'être associé à des changements d'utilisation des terres. La sélection du point de rupture associé à ce changement est basée sur l'hypothèse que la plupart des changements anthropiques induisent un changement dans le motif spectro-temporel de NDVI du pixel concerné. La détection des changements saisonniers est ainsi réalisée au niveau du pixel, à partir de séries temporelles du produit NDVI de MODIS dont la résolution temporelle (16 jours) et spatiale (250 m) est optimale pour la détection des IAGEs.

(2) Comme les changements détectés ne sont pas exclusivement liés aux IAGEs, la deuxième étape, toujours au niveau du pixel, vise à mieux comprendre les relations entre les principaux facteurs du changement et les types de changement (tels que révélés dans les séries temporelles, i.e., graduels, abrupts, saisonniers...), et à trouver un ensemble de métriques de changement discriminatives pouvant aider à identifier les changements d'usage des terres liés aux IAGEs.

(3) La dernière étape vise à affiner la discrimination et la caractérisation des IAGEs en incorporant des informations contextuelles à l'aide d'images satellites à plus haute résolution spatiale (Landsat).

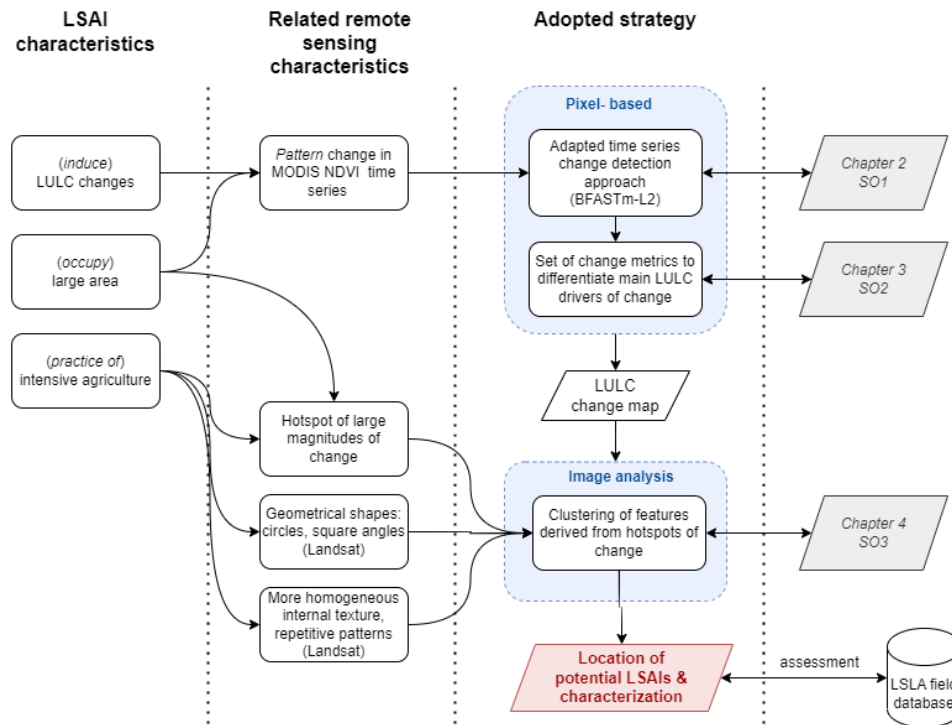


Figure 2. Cadre conceptuel de la thèse et structure du manuscrit.

L'objectif de recherche principal a été divisé en trois sous-objectifs interdépendants:

- Sous-objectif 1 : Détection automatique (non supervisée) des hotspots de changements anthropiques potentiels à l'aide d'un algorithme de détection de changement et de séries temporelles satellites à résolution moyenne (MODIS 250 m) appliqué à l'échelle nationale.
- Sous-objectif 2 : Affiner la détection des IAGEs en identifiant et spatialisant les facteurs de changements d'usage des terres : climatiques, naturels, anthropiques non agricoles et agricoles (incluant les IAGEs)
- Sous-objectif 3 : Identifier et caractériser les hotspots de changements d'utilisation des terres liés aux IAGEs à différentes échelles.

La thèse est structurée en cinq chapitres distincts. Après cette introduction, les chapitres 2, 3 et 4 présentent les résultats obtenus pour chaque sous-objectif (sous la forme d'un article scientifique), et le chapitre 5 expose les conclusions majeures et les implications et perspectives des résultats de la thèse.

## DÉTECTION NON SUPERVISÉE DE CHANGEMENTS D'UTILISATION DES TERRES BASÉS SUR LES CHANGEMENTS SAISONNIERS : BFASTM-L2

Ngadi Scarpetta, Y., Lebourgeois, V., Laques, A.-E., Dieye, M., Bourgoïn, J., Bégué, A., 2023. BFASTm-L2, an unsupervised LULCC detection based on seasonal change detection – An application to large-scale land acquisitions in Senegal. *International Journal of Applied Earth Observation and Geoinformation* 121, 103379.

Dans cet article (chapitre 2), nous proposons une méthode simple pour sélectionner automatiquement dans des séries temporelles d'images de télédétection, le point de rupture lié au plus grand changement saisonnier. Cette approche - nommée BFASTm-L2 - repose sur l'association d'un algorithme rapide (BFAST monitor) avec une métrique de similarité de séries temporelles (distance euclidienne L2) sensible aux changements saisonniers. La capacité de BFASTm-L2 à identifier la date de changement dans différentes situations a été testée sur deux jeux de données et comparée aux performances de trois autres algorithmes de détection de changement (BFAST monitor, BFAST Lite et Edyn). Ces jeux de données sont (i) un jeu de données de référence publié (Awty-Carroll et al., 2019), composé de 25 200 SITS simulés avec différents types (longueur de la saison (LOS), nombre de saisons (NOS), tendance, changement abrupt) et amplitudes de changement, et (ii) les STIS MODIS de NDVI de 2000 à 2020 sur une zone de 200 x 200 pixels au Sénégal comprenant différents sites d'étude ayant subi des changements d'utilisation des terres liés à des acquisitions de terres à grande échelle pendant la période étudiée, et rapportés dans la base de données terrain utilisée dans cette étude.

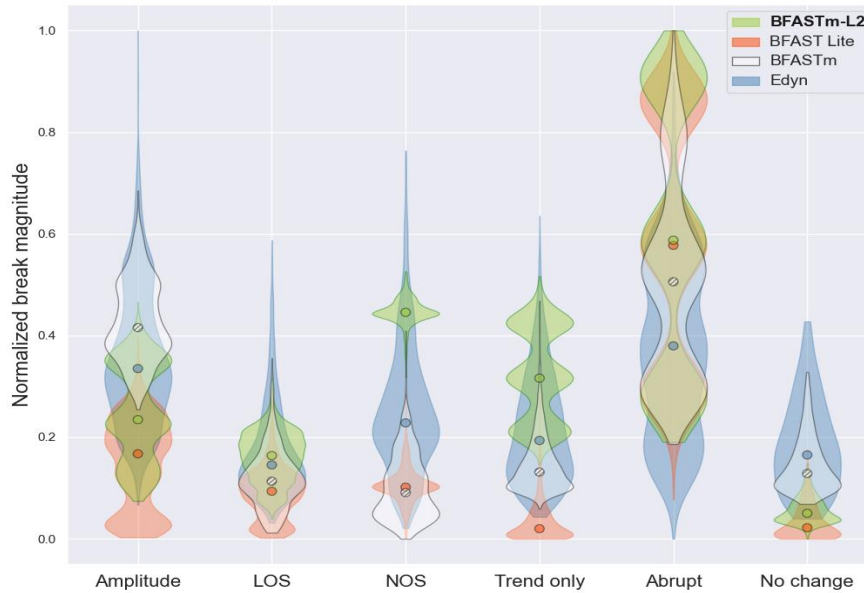


Figure 3. Diagrammes en violon de la plus forte magnitude normalisée pour chaque point de rupture, par type de changement pour : BFASTm-L2 (vert), BFAST Lite (orange), BFASTmonitor (blanc) et Edyn (bleu). Les points représentent la moyenne de chaque distribution.

Les résultats obtenus à partir des jeux de données simulés montrent que BFASTm-L2 est efficace pour détecter dans le temps la plupart des changements saisonniers qui sont caractéristiques des changements d'utilisation des terres liés à des activités agricoles (les moyennes des distributions de BFASTm-L2 sont les plus fortes pour deux (LOS et NOS) des trois types de changements saisonniers (Figure 3) en lien avec une activité agricole). Les résultats obtenus à partir des jeux de données réels montrent que, contrairement à BFAST Lite et BFAST monitor, les changements induits par les IAGEs sont spatialement mieux discriminés par BFASTm-L2.

L'algorithme proposé, plus rapide que BFAST Lite et Edyn, et avec une paramétrisation légère, peut donc être facilement mis en œuvre dans des pipelines non supervisés pour cartographier et analyser les changements génériques d'occupation et d'utilisation des sols à l'échelle nationale.

## CARTOGRAPHIE DES TYPES DE CHANGEMENTS PAR ANALYSE DES POINTS DE RUPTURE DANS LES SERIES TEMPORELLES D'IMAGES

Ngadi Scarpetta, Y., Lebourgeois, V., Laques, A.-E., Dieye, M., Bégué, A., 2024. Insight into large-scale LULC changes and their drivers through breakpoint characterization – An application to Senegal. *International Journal of Applied Earth Observation and Geoinformation* (submitted)

Cet article (chapitre 3) vise à apporter un éclairage sur les facteurs de changement des terres au Sénégal, via une carte composite RVB basée sur les changements détectés par BFASTm-L2 et caractérisés par trois métriques de changement dérivées des séries temporelles de NDVI MODIS. Ces métriques, choisies pour discriminer différents types de changements, sont la magnitude du changement, la direction du changement (ratio de NDVI) et la dissimilarité de la forme des séries temporelles.

La sensibilité de chaque métrique à différents types de changement a d'abord été testée sur jeu de données simulées (Awty-Carroll et al., 2019). Les mêmes métriques ont ensuite été extraites à partir de STIS NDVI MODIS (2000-2021) sur le Sénégal. La carte de changement RVB (Figure 4) a permis la visualisation de différentes "signatures" de changement, qui, combinées aux informations terrain, aux données de précipitations, à l'analyse des séries temporelles NDVI de pixels pour lesquels les changements sont connus et aux images Google Earth, ont contribué à les relier à différents facteurs de changement (Figure 5). Ainsi, ont pu être discriminés visuellement les changements liés à des facteurs climatiques de ceux liés à des facteurs anthropiques, tels que ceux induits par les investissements à grande échelle dans les systèmes agricoles intensifs ou l'exploitation minière



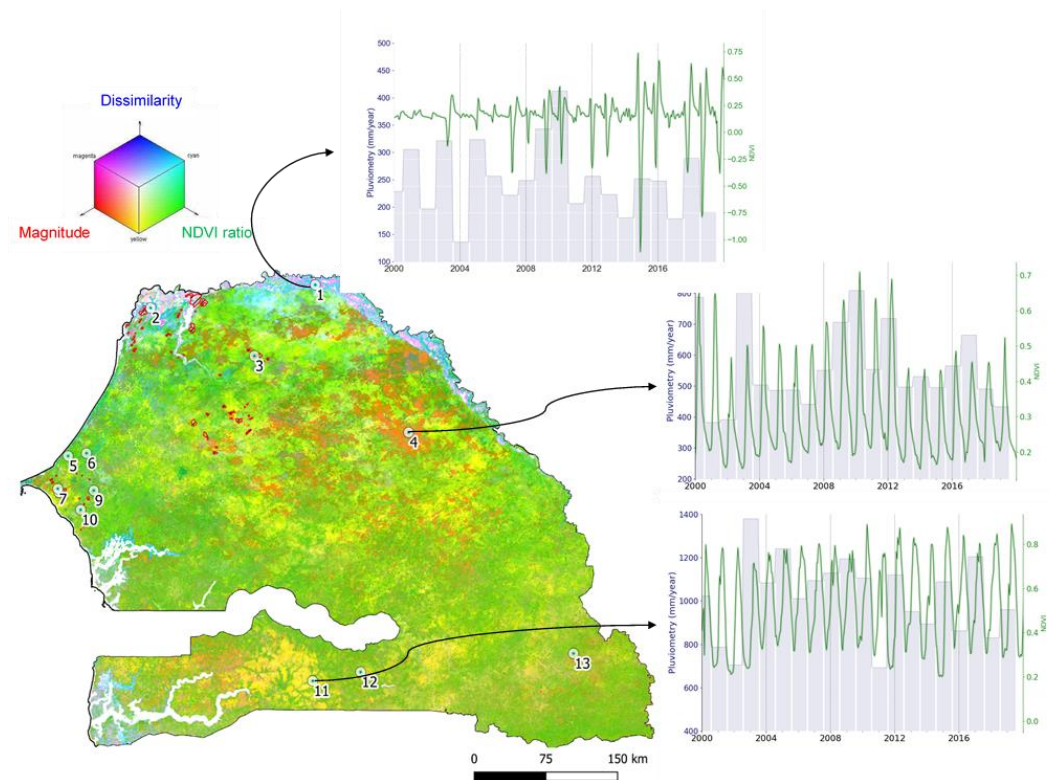


Figure 4. Carte composite RVB des métriques de changement avec en rouge : la magnitude du changement, en vert : le ratio NDVI et en bleu : la métrique de dissimilarité (valeurs étalées entre le 1<sup>er</sup> et le 99<sup>ème</sup> percentile). Les distributions annuelles de NDVI MODIS 2000-2019 et de précipitations TRMM sont présentées pour trois pixels d'écosystèmes différents : 1 (zone humide), 4 (savane arbustive), 11 (forêt tropicale sèche).

Band			Possible driver of change
R:	G:	B:	
Magnitude	NDVI ratio	Dissimilarity	
-	↓ = +	+	CLIM
-	↓ = +	-	NAT, CLIM
+	+	-	NAT, LSAI
+	+	+	LSAI
+	=	+	LSAI, CLIM
+	-	-	MIN/INFR, CLIM, (LSAIs in tropical ecoregions)
+	-	+	MIN/INFR

Figure 5. Tableau d'association entre les couleurs de la carte composite RVB (Figure 4) et les potentiels facteurs de changement (CLIM pour variabilité climatique ; NAT pour les changements biotiques naturels ; MIN/INF pour les mines/infrastructures ; LSAI pour les systèmes agricoles intensifs ou IAGEs).

# CARTOGRAPHIE ET CARACTÉRISATION DES INVESTISSEMENTS AGRICOLES À GRANDE ÉCHELLE AU SÉNÉGAL

Ngadi Scarpetta, Y., Lebourgeois, V., Laques, A.-E., Dieye, M., Bégué, A.. Mapping and characterization of large-scale agricultural investments in Senegal. *In preparation*.

Ce dernier article (chapitre 4) repose sur les résultats obtenus dans les chapitres précédents et vise à affiner la détection d' IAGEs au Sénégal par extraction (segmentation) des *hotspots* de changement, et classification et interprétation de ces hotspots. La méthodologie implique un processus en trois étapes, appliqué de manière indépendante aux régions des Niayes et de la vallée du fleuve Sénégal (SR pour *Senegal River*) :

1. Les changements d'occupation et d'usage des terres sont détectés à l'échelle nationale en utilisant l'algorithme non-supervisé BFASTm-L2 appliqué aux données NDVI MODIS. La magnitude du changement est ensuite pondérée en utilisant la dissimilarité de forme des séries temporelles et le ratio de NDVI pour mettre en évidence les IAGE potentiels.
2. Une analyse de contour est utilisée pour extraire les hotspots de changement, et des images Google Earth et des données de terrain d'IAGEs sont mobilisées pour identifier le facteur le plus probable du changement d'occupation et d'utilisation des terres, ainsi que pour labelliser les objets extraits servir de base de donnée de validation dans une étape ultérieure.
3. Pour chaque hotspot, des métriques spectro-temporelles (issues des STIS de NDVI MODIS) sont extraites à l'échelle de l'objet et comparées par des métriques spectrales (NDBal), texturales (indices d'Haralick) et structurelles (liées aux lignes détectées et aux formes rectangulaires/circulaires) issues d'images à haute résolution spatiale Landsat. Leur efficacité dans la discrimination des IAGEs est étudiée à l'aide de méthodes d'analyse non supervisée et évaluée à l'aide de la base de donnée de validation créée en point 2. .

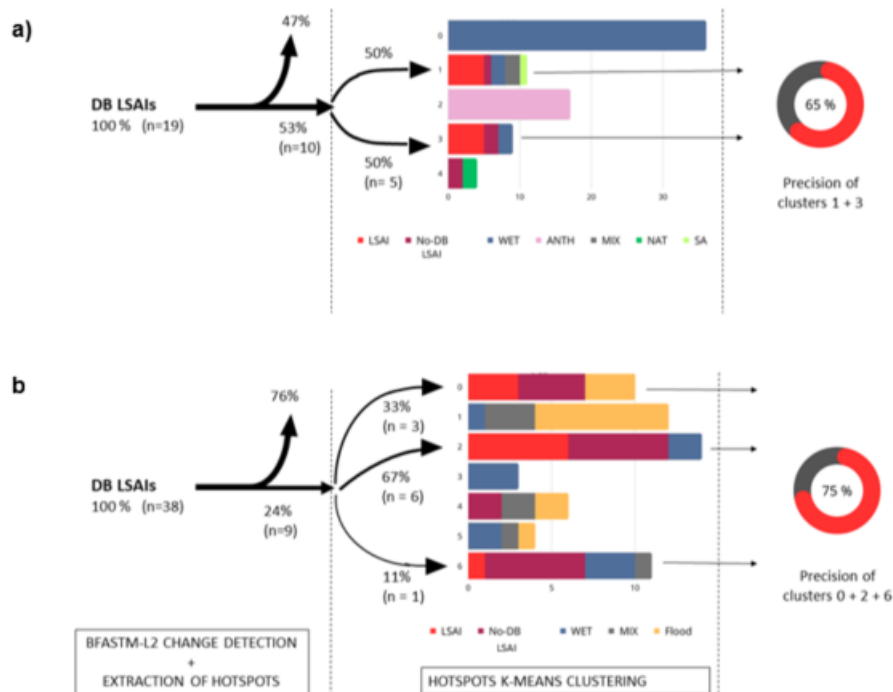


Figure 6. Pourcentage de détection des IAGEs rapportés dans la base de données de terrain (DB LSAIs) pour a) la région de Niayes et b) la région du SR. Les flèches verticales indiquent le pourcentage d'IAGE non détectés (pas de recouvrement entre les hotspots et les IAGEs de la BD). Les IAGEs détectés (flèches horizontales) sont lors de l'analyse non supervisée répartis dans 2-3 clusters et identifiés en tant que LSAI (rouge) et Non-DB LSAI (bordeaux, pour ceux n'apparaissant pas dans la DB). Les autres facteurs de changement identifiés comprennent : SA (petite agriculture, vert fluo), WET (zones humides, bleu), FLOOD (plaines d'inondation, orange), NAT (changements environnementaux autres que ceux liés aux zones humides, vert), ANTH pour mines, infrastructures et urbanisation (rose) et MIX (classe mixte, gris).

Les résultats montrent que les hotspots de changement issus de la segmentation d'image recourent 53 % (dans les Niayes) et 24 % (dans le SR) des IAGEs rapportés dans la base de données de validation (Figure 6). Le faible taux d'IAGEs détectés dans le SR est à relier à la relative petite taille des investissements (16% des investissements font moins de 50 ha, comparés à 5% dans les Niayes) et à la présence de zones humides et de plaines d'inondation dont la dynamique provoque des changements du même ordre ou supérieurs à ceux des IAGEs. Une analyse non supervisée des hotspots montre que les IAGEs dans chaque région peuvent être distingués des autres dynamiques de surface par leurs caractéristiques spectro-temporelles et structurelles. Les résultats d'un clustering de K-means basé uniquement sur les caractéristiques spectro-temporelles montrent que les taux de précision des clusters recombinaisonnés, comprenant les clusters les plus pertinents pour les IAGEs, sont de 65 % (dans les Niayes) et 75 % (dans le SR). Notons que les tests effectués dans différentes régions ont révélé que les caractéristiques structurelles ont un pouvoir discriminatoire variable selon la région analysée. Étant donné que toutes les IAGEs ne présentent pas des formes

géométriques distinctives (comme illustré dans l'exemple des plantations de caoutchouc au Laos dans l'annexe B.1), ces caractéristiques ont été exclues de l'analyse de classification pour assurer la robustesse et généralité méthodologique. Par ailleurs, les caractéristiques texturales ont également été exclues de la classification non supervisée en raison de leur faible pouvoir discriminatoire.

## DISCUSSION ET CONCLUSIONS

Ce travail de recherche méthodologique permet de répondre au besoin de suivi automatisé de l'utilisation des terres à grande échelle pour une gestion plus efficace des ressources et une prise de décision éclairée. L'accent est mis sur la détection des Investissements Agricoles à Grande Échelle (IAGEs) dans les zones tropicales, en raison de leur importance stratégique pour le développement territorial et de leurs impacts socio-économiques et environnementaux. Les IAGEs sont des systèmes complexes, avec une grande diversité de pratiques agricoles et d'environnements d'implantation, qui demeurent insuffisamment documentés à l'échelle nationale. Face à la complexité de collecte des données de terrain sur ce sujet sensible, l'effort principal s'est orienté vers le développement d'une méthode automatique, générique et interprétable pour détecter les IAGEs à échelle nationale.

Pour répondre à la question de recherche « Peut-on détecter automatiquement l'émergence d'investissements agricoles à grande échelle sans données de référence (c'est-à-dire de manière non supervisée) ? », un modèle conceptuel a été établi, reliant des métriques de télédétection (spectro-temporelles et spatiales) aux caractéristiques des systèmes d'utilisation des terres étudiés, et une stratégie en trois points a été proposée :

- Le développement d'un algorithme de détection de changement non supervisé, qui se concentre sur la détection automatique de changements significatifs liés à des activités anthropiques, principalement saisonniers, dans les séries temporelles NDVI de MODIS à 250 m.
- L'identification de trois métriques spectro-temporelles et leur combinaison dans une carte composite RGB, afin de relier les signatures et les changements d'utilisation des terres susceptibles d'être induits par des facteurs climatiques, naturels, anthropogènes (non agricoles) et agricoles (dont les IAGEs) à l'échelle nationale.
- L'extraction, l'analyse et la classification non supervisée de zones de changement potentiellement liées aux IAGEs. Ce point intègre des méthodes et des résultats des deux sous-objectifs précédents, en utilisant principalement des caractéristiques spectro-temporelles (issues des STIS de NDVI MODIS) et structurelles (issues de Landsat).

Ces trois points sont alignés sur le cadre proposé par Zhu et al. (2022) pour caractériser les changements dans l'utilisation des terres à partir de cinq facettes : *Où* (l'emplacement du changement), *Quand* (la date du changement), *Quoi* (ce qui change), *Comment* (les métriques du changement) et *Pourquoi* (le moteur du changement). La discussion résume les principales conclusions des travaux de thèse à travers le prisme de ces facettes du changement, en commençant par *Où* et *Quand*, puis en explorant les facettes *Comment* et *Pourquoi* qui en découlent.

## Où et Quand ?

L'algorithme BFASTm-L2 est un outil de détection de changements rapide et non supervisé, axé sur l'identification des points de rupture associés aux changements saisonniers significatifs dans les séries temporelles NDVI de MODIS. Cet outil comble principalement deux lacunes identifiées dans les algorithmes existants, à savoir : 1) l'ambiguïté quant au type de changement détecté par les algorithmes plus rapides qui ne procèdent pas à la décomposition des séries temporelles et 2) l'évaluation de la magnitude du changement, souvent basée sur les déviations par rapport à la moyenne, non adaptée à la détection et à la quantification de changements saisonniers. Les comparaisons avec d'autres algorithmes confirment la capacité de BFASTm-L2 à mettre en évidence les changements saisonniers associés aux investissements à grande échelle.

## Comment et Pourquoi ?

Pour répondre à ces questions, deux produits ont été développés :

- **La carte composite RVB des métriques de changement**, un outil de visualisation utile à l'échelle nationale : La carte nationale des changements d'occupation et d'usage des sols produite est basée sur les changements détectés par BFASTm-L2. Cette carte seule ne permet pas de distinguer exclusivement les changements dus aux IAGEs. Pour identifier les influences climatiques, naturelles, anthropiques et agricoles, trois métriques spectro-temporelles (la magnitude de changement, la dissimilarité de forme dans les séries temporelles et le ratio NDVI) sont calculées pour le plus grand changement détecté par BFASTm-L2 et combinées en une carte RVB permettant de caractériser les changements détectés (abrupt, graduel, saisonnier) et de les relier aux principaux facteurs de changements. Les résultats montrent que les IAGEs induisent principalement des changements saisonniers, confirmant les hypothèses sur les types de changements associés à différents moteurs.
- **Des cartes sous-nationales montrant les hotspots de changement détectés de façon automatique** à partir de la carte de magnitude des changements (détectés

par BFASTm-L2 et pondérée par la métrique de dissimilarité), ainsi que les clusters résultants d'une analyse non supervisée. Bien que l'intervention de l'utilisateur soit toujours nécessaire pour l'étiquetage des clusters, les résultats démontrent la faisabilité de la détection non supervisée des IAGEs. L'analyse des hotspots révèle quand à elle des caractéristiques distinctes pour les IAGEs, celles-ci se regroupant généralement en 2-3 clusters.

En conclusion, ces travaux ont permis le développement d'un algorithme efficace (BFASTm-L2) pour détecter les changements d'usage des terres, de caractériser ces changements en fonction des moteurs, et à distinguer les IAGEs des autres dynamiques terrestres. Les résultats contribuent à la compréhension des changements environnementaux à différentes échelles spatiales.

### Limites et perspectives

Les conditions requises pour une détection efficace des IAGEs avec BFASTm-L2 et MODIS sont les suivantes : 1) les objets cibles doivent présenter des surfaces de changement supérieures à 50 hectares ou être situés à distance des écosystèmes instables, 2) les IAGEs doivent impliquer des cultures à croissance rapide, et 3) l'approche doit être appliquée à des écorégions spécifiques. Lors de l'éventuelle extension de cette approche à des données provenant de satellites autres que MODIS, l'efficacité de BFASTm-L2 repose sur certains critères, notamment : 1) une fréquence temporelle élevée suffisante pour une représentation phénologique précise, 2) une durée minimale de 8 ans, et 3) l'absence de « trous » dans les données, avec un lissage visant à minimiser les fausses détections.

En ce qui concerne les perspectives de recherche visant à améliorer les produits, plusieurs recommandations sont avancées. Il s'agit notamment d'améliorer la vitesse d'exécution de BFASTm-L2, d'explorer de nouvelles métriques spectro-temporelles, et de conduire une analyse spatiale à l'échelle des objets pour une distinction plus précise des IAGEs.

Enfin, sur le plan opérationnel, la thèse vise à combler les lacunes d'informations sur le foncier, notamment dans les pays du Sud global, en soutenant la transparence de la gouvernance foncière et la prise de décisions. L'approche proposée, bien que ne fournissant pas une estimation complète de la superficie des investissements agricoles à grande échelle, représente une avancée cruciale. Des outils opérationnels et des programmes de formation pour accompagner cette démarche sont en cours de développement pour faciliter l'adoption de la méthode et accompagner les campagnes de terrain pour évaluer les résultats.

# 1 INTRODUCTION

## 1.1 WHAT ARE LARGE-SCALE AGRICULTURAL INVESTMENTS (LSAIs)?

### 1.1.1 The Large Scale Land Acquisitions (LSLAs), a global process

For more than two decades, there has been an intense debate in the media and literature about the number, evidence and distribution of **Large-Scale Land Acquisitions (LSLAs)**, more pejoratively known as "land grabs" (Batterbury, S., & Ndi, F., 2018; Borras et al., 2022; Bourgoin et al., 2019; Cotula, 2012; Davis et al., 2014; Schoneveld, 2014). This broadly refers to the occupation of large tracts of land by individuals, states or corporations (domestic or foreign), mainly for agricultural purposes (food crops, fodder crops, biofuels), but also for logging, tourism, conservation, mining, urban expansion or major infrastructure works. Although the phenomenon is not new, since the late 2000s there has been a growing global rush for land, driven by the confluence of increasingly volatile commodity markets, a rapidly growing global population, changing consumption patterns and climate change (see Figure 1.1).

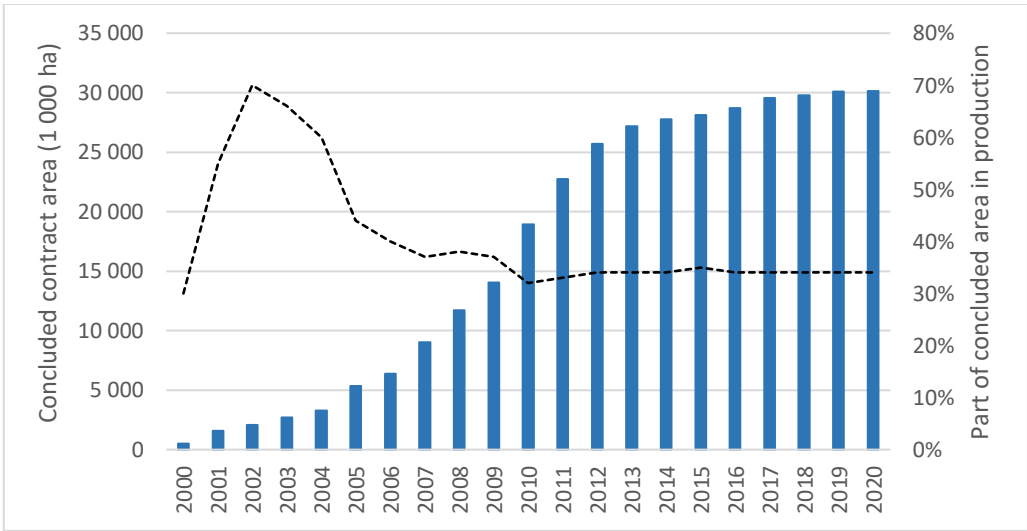


Figure 1.1: Cumulative global contract size under production (left axis, bars) and share of concluded deals size under production (right axis, dashed line)(adapted from Lay et al. (2021))

## Chapter 1: What are large-scale agricultural investments (LSAIs)?

The reported figures show a high geographical concentration of investments, with most of the rush for arable land concentrated in the global South and particularly in sub-Saharan Africa, which has become one of the main targets for LSLAs for plantation agriculture and forestry (Nolte et al., 2016; Schoneveld, 2014) (see Figure 1.2). Key target countries in Africa include (situation in 2016): Sudan, Mozambique, Ethiopia and Ghana. Interest in this region can be explained by the availability of favourable biophysical resources, low land and labour costs, the absence of strong land tenure systems and a strong government commitment to the development of commercial agriculture (Bourgoin et al., 2019; Schoneveld, 2014).

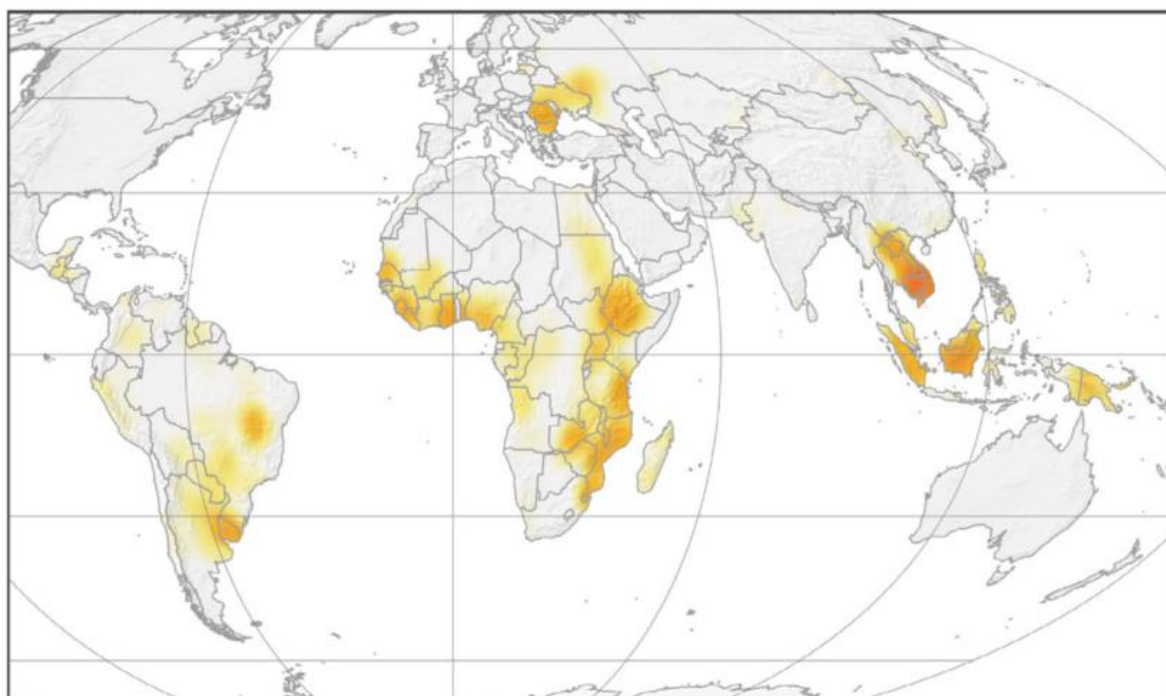


Figure 1.2: Global heat map of agricultural LSLAs contained in the Land Matrix in 2016. High densities of 943 concluded LSLAs are shown in orange, transitioning into yellow for lower densities. Projection: Mollweide. Source: M. Abebe (Nolte et al., 2016).

**The LSLAs for agriculture, here referred to as LSAIs (Large-Scale Agricultural Investments ; Box 1.1.1), represent the vast majority accounting for about 83% of concluded deals (Nolte et al., 2016).** According to the analysis of the global Land Matrix dataset <sup>1</sup>, an independent land monitoring initiative launched in 2012 in response to the thick opacity of related data, these LSAIs would account for 26.7 million hectares (represented by 422 deals) globally in 2016. Of these hectares, 37% would be located in Africa, representing approximately 42% of all concluded deals (Nolte et al., 2016). These figures, which differ

---

<sup>1</sup> <https://landmatrix.org/>



## Chapter 1: What are large-scale agricultural investments (LSAIs)?

significantly from various estimates found in the literature (Bourgoin et al., 2019; Schoneveld, 2014), only include *concluded* deals (leaving out 'intended' and 'failed' deals).

Schoneveld (2014) provided a detailed insight into the 'anatomy' of LSAIs in sub-Saharan Africa for the period 2008-2011. His study identified 353 projects larger than 2000 ha, covering about 8% of the region's annual harvested area. The vast majority (around 87%) had foreign investors (mostly from the EU and Asia), and most LAIs (63%) were for the production of biofuel feedstocks, particularly jatropha (around half of biofuel projects). However, as most of these rarely achieved the expected returns, many projects went bankrupt, temporarily suspended operations, scaled down or switched to other crops, **demonstrating the high spatio-temporal dynamics of these land objects**. Overall, investment in food crops was found to be much more limited, with sugarcane and rice being the main targets and other southern economies being the main investors (i.e. Asia and the Arab world, while North America and the EU dominated biofuel projects).

In terms of landscape characteristics, LSAIs typically cover much larger areas than average farms in the region (the size threshold used to distinguish them from smallholders can be as low as 20 ha locally and is often set at 200 ha for global analysis (Bourgoin et al., 2019)). They tend to use high levels of inputs (technology, agrochemicals, mechanisation...) compared to smallholders, leading to intensive farming practices, and often invest in irrigation systems and monocultures (Batterbury, S., & Ndi, F., 2018; Bey et al., 2020; Kuemmerle et al., 2013; Stefanski et al., 2014). As a result, and because of the larger field sizes, LSAIs landscapes are often less complex, more homogeneous, and less fragmented than in smallholder agriculture (Vogels et al., 2019).

### **Box 1.1.1- On the terminology used in this study**

This study focuses on the detection of LSLA for agricultural purposes using satellite remote sensing data. Although the primary focus of this study is on agricultural LSLA, the term **LSAI for "Large-Scale Agricultural Investment"**, or the terms **"Intensive Agricultural Systems"** and **"agribusinesses"**, which are closely related to agricultural LSLA, are used interchangeably in this study. The use of the term LSAI allows emphasis on the type of production system rather than on the land tenure which is not directly observable by remote sensing.

### 1.1.2 Importance of monitoring Large-Scale Agricultural Investments (LSAI)

Due to the growing importance of the LSAIs worldwide, many studies have focused on assessing their socio-economic and environmental impacts, which could be broader and more significant than previously thought, especially if 'non-operational' land deals are included in the analysis (Borras et al., 2022).

As the total area of alienated land in some countries can be huge relative to the amount of available arable land (e.g. Ghana: 61.6%, Ethiopia: 42.9% (Schoneveld, 2011)), these investments may undermine food security, with implications for rural livelihoods (e.g. disturbing social changes) and the environment (D'Odorico et al., 2017; Lay et al., 2021). Indeed, most of the land would come from customary or protected areas, with in some countries the impossibility of reverting to the previous status, leading to a long-term concentration of land resources and dispossession of vital livelihood resources (Nolte et al., 2016; Schoneveld, 2014). These impacts are all the more important given that markets are generally export-oriented, products are predominantly non-food, and the share of domestic investors is small, making the contribution of LSAIs to domestic market needs unlikely to be significant (Schoneveld, 2014). Alongside these negative impacts, some authors have highlighted the positive, albeit minimal (Davis et al., 2014) employment opportunities that some land investments can generate. Overall, the paucity of accurate data on the size, scope and distribution of land investments, particularly in sub-Saharan Africa, makes it difficult to assess their socio-economic and environmental impacts (Bourgoin et al., 2019; Schoneveld, 2014). Since official data are difficult to obtain (for a variety of reasons, including non-existence, limited access to data, decentralized administrations, and lack of consolidated data), most information comes from media reports, which carry a risk of bias, crowd-sourcing, and organizations with different working methods and definitions (what types of objects are actually included? For what period? Are planned/under negotiation/failed projects included?). All this has led to large discrepancies in quantifying the size and distribution of large-scale agricultural investments (Bourgoin et al., 2019; Cotula, 2012; Nolte et al., 2016; Schoneveld, 2014), as illustrated by the Senegalese case in Figure 1.3.

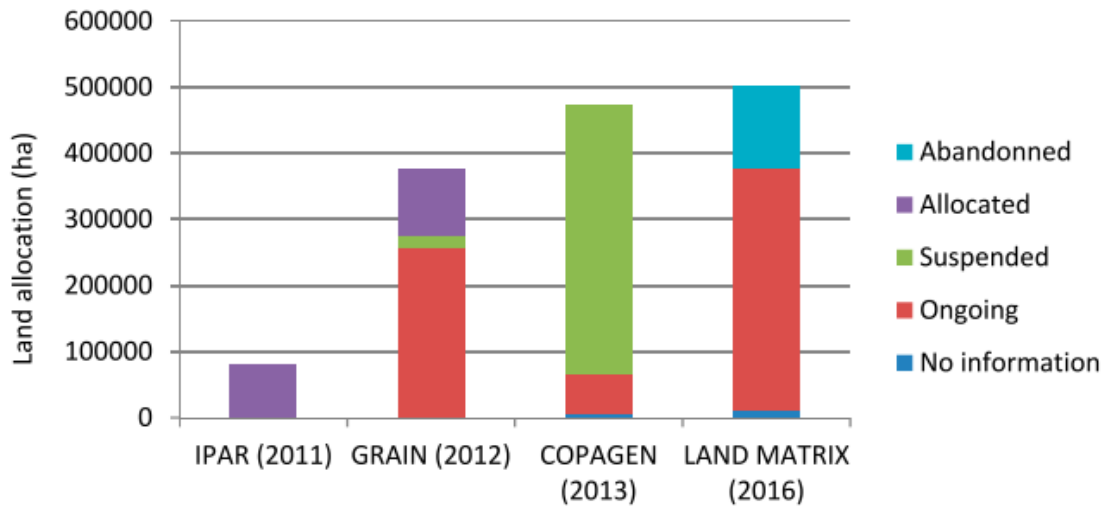


Figure 1.3: Status of reported LSLAs in Senegal from different sources (IPAR: Initiative Prospective Agricole et Rurale; COPAGEN: Coalition pour la Protection du Patrimoine Génétique Africain, GRAIN : international non profit organisation that support small farmers ). Source : (Bourgoin et al., 2019)

In Senegal, NGOs have decided to tackle the problem of the opacity surrounding LSLAs, by conducting two national inventories in order to quantify the LSLAs dynamics (Faye et al., 2011; Sy et al., 2015). Unfortunately, where inventories exist, they are often not up-to-date, incomplete and contain limited information (e.g. only point data, unknown implementation status, unknown type of crop) (Bourgoin et al. 2019; Schoneveld 2014). This is also observed in the Land Matrix global dataset, where accurate geospatial information is mostly missing or is not precise enough and is provided at an administrative level (region, village name, etc.). **There is therefore an urgent need for systems that can automatically detect, characterize and monitor large-scale agricultural investments in space and time. This requires unsupervised (to be applicable in regions where training field data doesn't exist or is scarce), tuning free (for genericity, transferability and operability / ease of use) and fast approaches (to be applicable at national scale).**

The ever-increasing availability of remote sensors, which have the advantages of monitoring the Earth's surface at repeated short intervals, providing consistent image quality, and being available free of charge, represents an interesting tool for the detection and characterization of LSAIs. The next chapter provides an overview of some satellite-based methods for the detection of LSAIs and related characteristics.

## 1.2 REVIEW OF EXISTING SATELLITE-BASED METHODS TO DETECT LSAs AND RELATED FEATURES

The signal detected by satellites is the result of the interaction between an electromagnetic wave and the constituent elements of a surface, such as vegetation and soil. Because remote sensing measurements are closely related to the biophysical and chemical properties of surfaces, these properties tend to be consistent for the same type of land cover in a given environment, be it forest, annual crop, surface water, bare soil, etc. Based on this premise, land cover maps are created using remotely sensed data. While land cover addresses the description of the land surface in terms of soil and vegetation, **land use mapping, which refers to the purpose for which humans use the land cover, does not have a one-to-one relationship with land cover, which makes it more difficult to capture.** In fact, a single land use can be associated with many types of land cover (and vice versa), as in the example of rice paddy, which represents an agricultural land use associated with a sequence of land covers throughout the year (i.e., vegetated land, barren soil, and flooded land) (Setiawan and Yoshino, 2014). As a result, **land use often requires thematic and/or local expertise and additional data** (e.g. socio-economic data) to interpret land cover classes, which in some cases may be based on observable activities (e.g., grazing) or structural elements in the landscape (e.g., pivot irrigation structures). Because the terms "land cover" and "land use" are closely related, they are often used interchangeably, and the commonly accepted term - LULC for Land Use Land Cover - recognizes that a mixture of the two is actually captured through remote sensing techniques. Even more challenging is the capture of land use *systems* (the coupled human-environment system) that describes how land, as an essential resource, is being used and managed. In the case of agricultural land use systems, this includes the ensemble of temporal and spatial cropping practices (Bégué et al., 2018).

LSAs represent a specific agricultural land use system that cannot be directly derived from remotely sensed imagery. They are complex land use systems with almost as many different forms as there are crops, environments, management techniques and practices. As a result, various strategies have been used to detect the LSAs, most of which are based on **the detection of agricultural intensification practices**, through a set of satellite-derived spatio-temporal proxies (Kuemmerle et al., 2013):

- (i) Cropping intensity, defined as the number of cropping cycles observed per year on a unit area of cropland (i.e., the number of peaks in a pixel's vegetation index time series) (Hentze et al., 2017);

- (ii) Irrigation (Ozdogan et al., 2010). Since irrigated agricultural land is expected to be more productive, temporal differences in biomass and greenness indices are exploited, especially during the dry season (Eckert et al., 2017). Besides the impact of irrigation on vegetation, some studies have also focused on the detection of irrigated structures (Tang et al., 2021);
- (iii) Field size (Graesser et al., 2018; Kuemmerle et al., 2009). While most of the spectral features used are derived from visible and near infrared (NIR) bands (e.g. NDVI being the most common), Synthetic Aperture Radar (SAR) has also been successfully used to derive metrics related to the geometry and topology of croplands (Ajadi et al., 2021; Liu et al., 2019; Stefanski et al., 2014).

As the implementation of LSAI is often associated with changes in land use (i.e. the adoption of intensive agricultural practices (see Section 1.1.1)) and land cover (e.g. the conversion of natural vegetation to bare soil) at large spatial scales, some strategies have introduced **LULC change techniques** that aim to exploit the temporal dimension and add change characteristics to the analysis. This reduces the complexity of the analysis by focusing on first identifying areas of LULC change that may be associated with LSAs. Although many methods have been developed for detecting LULC change, many of which have been discussed in various reviews (Chughtai et al., 2021; Lu et al., 2004), the use of remote sensing to detect and analyse land use systems such as LSAs is still in its infancy (Verburg et al., 2019; Weiss et al., 2020).

With this in mind, and with the aim of better understanding the challenges and opportunities in detecting LSAs using remote sensing data, this subsection outlines the evolution of some satellite-based methods, starting with LULC classification techniques to detect intensive agricultural systems (1.2.1.), and progressing to monitoring LULC changes in a supervised (1.2.2.1) and unsupervised (1.2.2.2) manner.

### 1.2.1 Mapping intensive agricultural systems

The simplest, but most time intensive and expensive method of mapping intensive agricultural systems has been the manual digitisation of satellite imagery (Degife et al., 2018) or computer-assisted photointerpretation (CAPI). Due to the difficulty of applying CAPI over large-scale areas, (semi-)automated approaches have been developed to assess agricultural intensification by **extracting agricultural fields and assessing their size and extent** (Graesser and Ramankutty, 2017; Kuemmerle et al., 2009; Yan and Roy, 2014). These are often image processing methods (combining edge-based methods, segmentation and image morphology) applied at a regional scale, either to predominantly agricultural regions (Yan and Roy, 2014) or to filtered pixels after applying a crop mask (Ajadi et al., 2021; Graesser and Ramankutty, 2017). Indeed, it is generally accepted that field size is positively correlated with

agricultural intensity (Degife et al., 2018; Kuemmerle et al., 2009; Vogels et al., 2019). These methods are therefore highly dependent on field size, shape, configuration and crop type between neighbouring fields to accurately identify field boundaries (Graesser and Ramankutty, 2017).

A different and very popular strategy that has been used to map specific cropping systems, is the use of supervised classification techniques that rely on the **separability of specific crop types based on their phenological patterns** and seasonal dynamics observed in satellite-derived vegetation indices (Arvor et al., 2011; Sedano et al., 2019). The number of crop cycles is often used as a proxy for agricultural intensification. For example, these approaches have allowed the mapping of specific major land-use systems, such as double-cropping systems of commercial soybean, maize and cotton crops over large areas in the USA (Wardlow et al., 2007) and Brazil (Arvor et al., 2012; Bellón et al., 2017; Chen et al., 2018). In Asia, many studies have been conducted to map specific boom crops such as rubber (Fan et al., 2015; Xiao et al., 2020). A different approach is the one developed by Sedano et al. (2019), who developed a LULC mapping framework specifically designed for agricultural systems of the Sudan-Sahel region, based on a Knowledge-Based Expert System (KBES) (i.e., a type of AI system that incorporates human knowledge often in the form of a set of rules ) on the region phenological cycles.

While very successful, pixel-based methods do not integrate spatial contextual information. Since intensive agricultural practices (e.g., mechanization, irrigation...) often result in **distinctive spatial arrangements and homogeneous landscape textures** (Kuemmerle et al., 2009; Özdoğan et al., 2018), the Object-Based Image Analysis (OBIA) (Blaschke, 2010) or GEographic OBIA (GEOBIA) approach has also been implemented as it takes advantage of these spatial features. It involves segmenting satellite images (generally having a high / very high spatial resolution) into homogeneous discrete objects or regions, and analyzing various attributes (eg. shape, texture, geometry, adjacency, etc) and spectro-temporal information to classify land use. For example, Vogels et al. (2019) applied a GEOBIA classification approach to map LSAs in a highly contrasting landscape in the Central Rift Valley of Ethiopia using high spatial resolution satellite imagery (SPOT6, 6m). In their study, texture and shape information appeared to be essential for the identification of LSAs. This correlated with previous studies showing a strong relationship between field size and texture metrics (Kuemmerle et al., 2009). While these approaches have advantages over the pixel-based approaches presented above in terms of capturing spatial patterns, GEOBIA in addition to requires temporal stability of the object for reliable classification results. Furthermore, since GEOBIA is not data-driven, it requires adjustment for each region of interest. Since this adjustment involves several

parameters (including segmentation, spatial, spectral...), the analysis is not directly applicable to regions with different LSA characteristics, limiting the genericity of the approach.

More recently, the increasing availability of free remote sensing data and the overwhelming increase in computing power have led to the development of **new machine and deep learning algorithms for image supervised classification**. Tang et al. (2021), were able to accurately map pivot irrigation systems in the state of Mato Grosso, Brazil, based on an object-based deep learning detection approach, combined with a deep learning image classifier capable of distinguishing pivots from other circular objects in Sentinel-2 images. Accurate object delineation was then achieved using a shape detector (Hough transform). Ajadi et al. (2021), based on optical and SAR data, used a combination of deep learning approach to extract large-scale field boundaries across Brazil, with a spatially adapted supervised classifier to map the crop type. While these methods are data-driven and very powerful in capturing spatial patterns, they are still underexploited in their ability to consider both spatial and temporal (historical data) contextual information (Molinier et al., 2021). Furthermore, an important challenge is to build a large high-quality training dataset that represents not only the spatial and temporal variability of the target, but also of the observation conditions, including the noise and biases to ensure their generalization capability (Weiss et al., 2020). However, as with GEOBIA, the transferability of these approaches to regions with different spatial characteristics is not guaranteed (e.g. regions with less contrasting structures between smallholder and large-scale agriculture).

In summary, while these approaches have been able to map some large areas of major land use systems at a given point in time with high accuracy, they are mostly supervised approaches that rely on large amounts of training data, are crop and region specific (Estel et al., 2016; Graesser et al., 2018), and overall are not easily transferable to other regions. Most importantly, these methods provide a **static representation of land cover at one point in time and are not designed to capture temporal changes in LULC, limiting the ability to analyze trends and transformations over time**. For this reason, and with the increasing availability of long stacks of satellite imagery, methods that can accurately capture land dynamics over time are sought and are the subject of the next section.

## 1.2.2 Monitoring LULC changes associated to LSAs

### 1.2.2.1 Supervised approaches

The methods presented so far have been shown to accurately map land use systems at a given point in time. However, they do not take advantage of the full temporal depth available in long and dense satellite imagery time series (SITS) and, in the case of the object-based

studies, they may miss detections due to selection of the timing of image acquisition (Tang et al., 2021). As a result, recent studies have focused on detecting LULC changes associated with LSAI, even at the pixel level. This shift in focus is significant as it allows for a more detailed and dynamic understanding of LSAI-driven transformations in the landscape.

Land change detection has been defined by Singh (1989) as "the process of identifying differences in the state of an object or phenomenon by observing it at different times". Among the various methods used, image differencing change detection and especially **post-classification change detection**, which compares LULC (often supervised) classification results at different times, have been most widely implemented (see examples of Chen et al. (2023a) and Özdoğan et al. (2018) for rubber plantation detection), although these methods, designed for one-time classification, are not suitable for monitoring changes over time, as errors in the initial classification are propagated to the results (Zhu, 2017). Transitioning from the detection of specific crops such as rubber crops to the detection of specific land use systems, Bey et al. (2020) used a random forest-based approach that integrates contextual (i.e. textural) features to identify land use trajectories between smallholdings and large-scale croplands in the Gurué district of Mozambique, assuming that field size and field homogeneity of LSAI are greater than those of smallholdings. Detection of change was achieved by comparing classified land use maps derived from three-year composite images. Similarly, Eckert et al. (2017) assessed land change associated with large-scale commercial agricultural systems in the foothills of Mt. Kenya using a post-classification method based on contextual information. The "irrigated agriculture" class was defined as "large-scale commercial agriculture" if greenhouses and water bodies covered more than 3% of a land unit (2 km x 2 km).

With the increasing availability of long and dense SITS, recent studies have shifted from post-classification approaches, which limit analysis to a few pairs of classified LULC maps, to **trajectory classification approaches of SITS**. These supervised approaches aim to classify different types of LULC change based on time series trajectories learned from a training dataset. According to Zhu (2017), these methods are mainly designed to detect abrupt changes. An example is the study by Hurni et al. (2017), who used long and dense SITS to train a support vector machine (SVM) supervised classification to map land transitions to specific 'boom' crops (rubber, cashew, eucalyptus, sugarcane and coffee) in Southeast Asia. At the global scale, Curtis et al. (2018) focused on drivers of deforestation and using decision-tree models mapped the contribution of 'commodities' (including agriculture with oil palm plantations, mining and energy infrastructure) to deforestation, assuming that commodity-led deforestation is characterised by permanent conversion of forest/shrubland to non-forest land use. More specifically, they distinguished commodity-driven deforestation from shifting



agricultural deforestation by temporally assessing forest/shrub regrowth after deforestation, using a land unit size of 10 x 10 km.

In addition to the commonly used post-classification and trajectory classification approaches, different methods have been tested, such as that introduced by Chen et al. (2023b) whose objective was to assess the main drivers of large-scale forest disturbance in Laos (among them: “new plantations” and “shifting agriculture”). In this study, change was first detected in the time series using **Continuous Change Detection and Classification - Spectral Mixture Analysis (CCDC-SMA)**. The category of interest (i.e. “New plantations”) was then identified using a post-disturbance classification combined with a thresholding on the magnitude of change.

As observed, many different strategies have been used to monitor LSAs with Earth Observation data, with varying degrees of accuracy and automation, gradually taking advantage of the increasing availability of new data. **So far, most relied heavily on the availability of reliable training data sets.** As will be discussed in more detail in section 1.3.1, **these data are difficult to obtain** for LSAs and, in most cases, far from perfect, which affects the accuracy and generalisation capability of the classifiers. For this reason, unsupervised approaches are sought. Progress and challenges in this area are presented in the next section.

#### *1.2.2.2 Unsupervised approaches*

Automatic approaches are sought that do not require training data, rely on a small number of parameters, and are computationally efficient to be scalable to large scales. Such studies using unsupervised techniques for LSAI detection are still scarce.

**To date, most of these unsupervised approaches have been applied to detect major land cover changes happening on forested ecosystems (Ochtyra et al., 2020), and few to detect changes associated with agricultural systems.** These studies, often crop-specific and using multi-temporal satellite data, have focused on detecting distinctive phenological patterns associated with specific agricultural land use systems. Among these, several studies were found that mapped the target land use system in two different time periods, and compared them for change detection. This is the case of Kontgis et al. (2015), who detected rice fields at two different time periods, first discriminating rice from other land cover types at the pixel level based on the variability of Vegetation Indices (VIs) spectral indices and defined hard thresholds. Then, after transforming the pixel-based map into an object-based map, a trajectory approach was used to discriminate rice objects with multiple cropping cycles per year - a proxy of agricultural intensification. Changes related to agricultural intensification were then detected by comparing maps between two different time periods. Despite the use of multi-temporal data, these approaches still capture crop fields at a single point in time, typically

relying on three-year imagery epoch data. This requires pixel stability, meaning that land use has not changed within the three-year timeframe of the imagery used for classification. In these approaches, multi-temporal data are often used to mitigate data gaps and reduce spatial and spectral ambiguity, as well as to extract temporal descriptive statistics related to specific phenologies. In addition, and as consequence of not using training datasets, they often require prior differentiation of cropland to perform their analysis.

Rather than looking at change through multi-date comparisons, recent studies have focused **on detecting change directly within full stacks of satellite imagery**. This is the case of (Ye et al., 2018), who used the presence of persistently low values in annual aggregated NDVI time series to detect rubber plantations (these values are characteristic of the implementation of new plantations; i.e. land clearing/planting preparation). Focusing on the characteristics of some specific land use systems rather than specific crops, Hentze et al. (2017), using a Seasonal Trend Analysis (STA) combined with the BFAST change detection algorithm detected changes in the number of crop cycles characteristic of changes in agricultural intensity (i.e. irrigation) in Zimbabwe. However, this again required prior crop masking with a third-party land cover classification product.

To summarize, although many unsupervised approaches to land use system detection avoid the need for a training dataset, they often require user intervention, rely on intermediate products such as crop masks, or are tailored to specific crops (and region specificities), making their application in different regions challenging.

### 1.2.3 Synthesis

As can be seen, methods for detecting intensive agricultural systems are becoming increasingly complex, ranging from LULC classification approaches representing land cover at a single point in time, to more complex approaches using extensive satellite imagery time series. Proposed approaches for detecting intensive agricultural systems are mostly based on: (1) distinctive phenological patterns over time (often based on the number of cropping cycles or related to the implementation of irrigation systems), (2) specificities of landscape objects in size and arrangement in space (field size, field geometry, homogeneity...). However, these existing methods are not without limitations.

## 1.3 IDENTIFIED CHALLENGES IN REMOTE SENSING TO DETECT LSAI RELATED CHANGES

Chapter 1.2 highlighted the limitations that characterize most LSAI detection approaches to date. These limitations can be categorized as:

- The limitations imposed by the spatial and temporal resolutions of available remotely sensed data, as well as the challenges posed by the presence of data gaps;
- The limitations related to the different underlying assumptions about what defines an LSAI (see study cases in Appendix A) ;
- The limitations related to the methodological choices, such as: (i) the chosen classification scheme, where the number of classes directly affects intraclass variability and interclass similarity (Hurni et al., 2017), (ii) the careful selection of appropriate spatial and temporal satellite imagery representing key phenological windows (Bey et al., 2020; Chen et al, 2023a; Vogels et al., 2019), (iii) the compositing method used to derive the texture metrics (if any) used in the classification (Bey et al., 2020), and (iv) the area of the cells used to extract contextual information (if integrated) (Curtis et al., 2018). All of these explain the different results that can be obtained for the same region, as shown in Box 1.3.1. In addition it worth highlighting that these studies often focus on a specific crop or regions dominated by a specific disturbance type, potentially overlooking other disturbance types (Li and Fox, 2011). Consequently, it remains uncertain whether these methods are effective when applied to larger regions with a variety of change types.

### **Box 1.3.1– The role of rubber plantations in deforestation in Laos**

Methodological choices and assumptions can lead to large differences in the results obtained. This is particularly true for supervised approaches, which can be very sensitive to many different parameters (see Section 1.3.). For example, Hurni et al. (2017), using a SVM classifier found that **7 360 km<sup>2</sup>** (or 3.6% of the study area consisting of seven Landsat footprints in Southeast Asia, four of which cover Laos and three of which are exclusively in Laos) were converted from forest to rubber plantations between 2000 and 2014. Although not directly comparable, a much lower figure was found by S. Chen et al. (2023) who, using a CCDC-based change detection algorithm combined with a post-disturbance classification on Landsat imagery , found that less than **1% (only 969 km<sup>2</sup>)** of forest disturbance in all the national territory of Laos between 1991 and 2020 was due to "new plantations", while the Land Matrix reported about **4 000 km<sup>2</sup>** dedicated to rubber alone in 2020 (Land Matrix, 2023).

In addition, important **limitations specific to supervised or unsupervised approaches have been identified**. The first one, specific to supervised approaches, is represented by the difficulty of obtaining a large and reliable ground truth dataset of LSAIs that ideally fully represents their spatio-temporal heterogeneity. This dataset must not only be large enough, but also well balanced, as class imbalance can lead to under/overclassification problems. In contrast, unsupervised approaches face a critical challenge in selecting an appropriate change detection algorithm and distinguishing detected changes of interest from other drivers of change.

These last three challenges, i.e. (i) the acquisition of representative ground truth data, (ii) the change detection algorithm, and (iii) the identification of change drivers, critical to the choice between a supervised/unsupervised approach, are discussed in more detail in the next subsections.

### 1.3.1 Difficulties in obtaining reliable training and validation data, and implications

As observed throughout the implementation of this work and evidenced by the Land Matrix Initiative, obtaining field data (with digitized land deals) on LSAIs is challenging. And when available, field datasets are often incomplete, not up-to-date and imperfect due to inaccuracies associated with survey interviews, the high spatio-temporal dynamics of LSAI objects and issues such as: (i) the reported date of change is often the date of the transaction and not the date of actual implementation, (ii) the reported areas under contract are often underused or abandoned, (iii) the areas under contract do not correspond to the areas in the field (Özdoğan et al., 2018). Because of the difficulty in obtaining ground truth data, many studies rely on visual interpretation from high-resolution satellite imagery (Bey et al., 2020; Chen et al., 2023a; Curtis et al., 2018; Özdoğan et al., 2018; Ye et al., 2021) to derive LSAIs. However, this method is not straightforward, since: (i) land use is not directly inferable from land cover, (ii) some crops/plantations are visually indistinguishable from their surroundings, making it difficult to identify boundaries (e.g. forestry deals embedded within forested areas), (iii) unavailability of high-quality images (sensor characteristics, gaps) suitable to accurately digitized LSAIs, (iv) infer the current status (still active?) may be very difficult, as illustrated by the results of the *Mapathon* coordinated by the Land Matrix in 2020 (see Box 1.3.2).

**Box 1.3.2– Insights from the 2020 Mapathon on the difficulty of digitizing land deals using satellite imagery**

In 2020, Land Matrix launched a mapathon with a group of 38 participants who received training and support from technical advisors. Their task was to delineate 1159 LSLAs boundaries, representing 20% of all transactions registered in the Land Matrix global database, using high-resolution Google Earth imagery (Bourgoin and Harding, 2021). Although these deals were initially classified as "highly accurate" in the database, meaning that they were geolocated (with a point) or associated with a village/hamlet name, only 34% of them had identifiable field boundaries that could be digitized. The remaining deals could not be digitized, mainly due to the inability to visually discern boundaries or because the transaction no longer existed. A very small fraction (2%) could not be reached due to image blur. This highlights the challenges of obtaining accurate geospatial information for land transactions and maintaining an up-to-date field database.

**This justifies the need for unsupervised approaches based on the detection of spectral and spatio-temporal features typically induced by changes in land use systems such as LSAIs.** The massive amount of temporal information carried by SITS, in addition to its temporal depth, makes SITS a promising material to work with.

### 1.3.2 Adequacy of existing time series unsupervised change detection algorithms

Unsupervised time series change detection algorithms play a crucial role in revealing patterns of land use and land cover (LULC) change, but their effectiveness is not without limitations. Primarily designed to detect abrupt changes (e.g. deforestation) and gradual changes (e.g. logging or afforestation), very often in forested areas (Ochtyra et al., 2020), **these algorithms face challenges in detecting more subtle changes that may characterize shifts in land use.** As shown in Section 1.2, newly established LSAIs, when active, most of the time induce changes in cropping practices (if the previous land use was already agricultural) and, if significant enough, in land cover. **These changes often manifest as pattern change in phenological time series** such as those represented by vegetation indices (Mardian et al., 2021; Setiawan and Yoshino, 2014). More particularly, these include changes of seasonal type such as the number of cycles (the basis of many of the studies focused on detecting intensive agriculture of Section 1.2), the amplitude of the growing cycle peak and the length of the growing cycle.

Detecting changes in time series is not an easy task, as periodic seasonal changes can mask sporadic yet meaningful changes. To mitigate the impact of these seasonal changes, various approaches have been developed. These include temporal aggregation of time series data, the use of specific time windows (e.g. growing season), and the normalisation of reflectance values per land cover type (Verbesselt et al., 2010a). Among unsupervised algorithms that use full temporal data, three main strategies are commonly employed to account for seasonality in time series:

- The segmentation-based algorithms (e.g. LandTrend (Kennedy et al., 2010), Continuous Change Detection and Classification (CCDC) (Zhu and Woodcock, 2014), Detecting Breakpoints and Estimating Segments in Trend (DBEST) (Jamali et al., 2015)), which divide the time series into distinct segments that show consistent behaviour (Jamali et al., 2015). Changes (usually abrupt and gradual) between these segments are then assessed.
- The seasonal-trend decomposition techniques such as STL (Seasonal-Trend decomposition using LOESS), which separates the time series into seasonal, trend and residual components, facilitating the identification of changes within each component (e.g. Breaks For Additive Seasonal and Trend (BFAST) (Verbesselt et al., 2010b)). Although effective for the detection of abrupt/gradual changes, this is a computationally expensive procedure less well performant in the detection of seasonal changes (Zhao et al. 2019).
- The algorithms that incorporate adaptive modelling to dynamically adjust for seasonality (e.g. BFAST monitor (Verbesselt et al., 2012), Exponentially Weighted Moving Average Change Detection (EWMACD) (Brooks et al., 2014)). These are often faster than those that perform seasonal trend decomposition and are often part of online change detection algorithms that, unlike offline algorithms such as those mentioned in the previous two paragraphs, are able to detect changes in the data as they become available over time in near-real time. Changes are usually detected in the residuals and correspond to significant deviations between actual observations and predicted values. Since larger deviations from the mean are caused by abrupt changes or changes in amplitude, these types of changes can be expected to be selected over more subtle seasonal ones.

**The analysis is further complicated by the use of long and real time series. Indeed, the longer the time series, the greater the likelihood of detecting a combination of several types of change, in which case selecting the most representative breakpoint becomes a non-trivial task.** This is even more true for LSAI induced changes, which, as mentioned above, very often induce seasonal changes, but also abrupt changes (e.g., due to increases

in biomass productivity or deforestation). Therefore, approaches that only respond to seasonal changes (as in (Hentze et al., 2017; Mardian et al., 2021) are not appropriate in this context.

Furthermore, accurately identifying the point at which the pattern changes the most does not guarantee that it is related to LSAI. **This requires a better understanding of the changes induced by different key drivers.** This challenge is outlined in the next section.

### 1.3.3 Discrimination of LSAIs induced changes from other drivers of change

Differentiating the main drivers of LULC change is essential for effective resource management and informed decision making. Although essential, distinguishing them from inter-annual climate-related variability and trends is not an easy task. To date, many unsupervised approaches have been developed, primarily relying on correlations between vegetation productivity and meteorological data (e.g. pluviometry, temperature...) (Anchang et al., 2019; Leroux et al., 2017; Xiao and Moody, 2005). One popular approach is the Residual Trend Analysis (RESTREND), designed to disentangle climatic and anthropogenic drivers of land degradation using temporally aggregated data (Evans and Geerken, 2004; Wessels et al., 2012). However, RESTREND has limitations tied to the assumption of a strong relationship between vegetation production and rainfall, making it suitable only under specific conditions (e.g. not too severe land degradation), and its effectiveness is influenced by the choice of the assessment period (Wessels et al., 2012). Based on the same assumption, Dutrieux et al. (2015), using the BFAST change detection algorithm, introduced an external precipitation-based regressor to their model to account for natural variability in a dry tropical forest area in Bolivia. However, contrary to expectations, this modification diminished the model's predictive power, showing that the relationship may be more complex than assumed.

Approaches based on vegetation productivity need to be treated with caution when considering the identification of LSAIs due to their high spatio-temporal dynamics and high variability in agricultural practices. Assuming that all LSAIs practice intensive agriculture, high values of NDVI as a proxy for Net Primary Production (NPP) cannot be universally relied upon, as they depend on the climatic condition, specific crop types used and are not exclusive to LSAIs. In fact, depending on the region studied, small farms may also use intensive practices (e.g. countries in the northern hemisphere). Therefore, with the availability of long and dense SITS, it may be more appropriate to focus on the pattern of change and change metrics to disentangle different drivers of change, rather than solely relying on NPP/NDVI trends. While this aspect of change is implicit in supervised change detection approaches, it remains underexplored in unsupervised approaches. Indeed, **it is generally accepted that within time series, long-term trends usually result from slow, gradual processes such as climate change or forest degradation, abrupt changes are due to environmental disturbances**

**to land cover such as fires, deforestations, or floods, and seasonal changes due to changes in vegetation phenology, composition or management practices** (Hentze et al., 2017; Mardian et al., 2021; Setiawan and Yoshino, 2014). However, a more in-depth study of this facet of change research could be beneficial, for example, to explore whether changes associated with climate variability or climate change are primarily manifested as changes in amplitude and trends, and to assess the likelihood of impacts on variables such as the number of growing cycles. **This finer analysis of the type of pattern change induced in the time series, combined with different change metrics derived from the time series and spatial contextual information, should help to differentiate the main drivers of LULC change.**

## 1.4 RESEARCH OBJECTIVE

### 1.4.1 Scope

While substantial research has been dedicated to the detection of land cover change, there remains a lot to be done to detect more complex land use (system) change (Verburg et al. 2019). LSAs are part of such land use systems that need to be monitored on a global scale for improved transparency, efficient resource management and informed decision-making. Current research on LSAs detection has predominantly focused on supervised approaches based on machine learning techniques and, more recently on deep learning techniques, given the increasing availability of high to very high spatial resolution data. However, these techniques rely on the availability of a very large, representative and difficult-to-obtain training dataset, and they are particularly sensitive to the conditions under which the dataset was acquired (i.e. geometry of acquisition, atmospheric conditions, crop conditions...). In addition, these approaches are often region- and/or crop-specific, and based on assumptions that may not be transferable to other regions. For all these reasons, **this study aims to develop an unsupervised and generic framework designed to detect LULC changes at large scales potentially associated with Large Scale Agricultural Investments (LSAIs), using satellite imagery and gathering evidence through identification of drivers of change and landscape analysis.**

LSAIs are defined here as compact farms that occupy at least 30 hectares of land used for intensive agriculture (i.e. mechanized), regardless of their land tenure (e.g. agribusinesses, cooperatives...). The scope of this study, while attempting to be as general as possible, focuses on Senegal as a case study and covers the period between 2003 and 2018. Senegal's diverse ecoregions, ranging from arid in the north to semi-humid in the south, provide a rich landscape for study. The country experiences two distinct climatic seasons: a dry period from November to May and a rainy season from June to October. Steppe, savanna and sub-humid dry forests



dominate the land cover, as documented by various researchers (Budde et al. (2004), Sultan and Janicot (2003; Tappan et al.) and Tappan et al. (2004)).

In addition, Senegal has a unique advantage with its National Land Observatory, supported by the Land Matrix, which contains a comprehensive database on large-scale land acquisitions (LSLAs). In 2019, the Senegalese Institute of Agricultural Research (ISRA) conducted a field campaign on LSLAs, recording more than 700 polygons in a spatial database (M. Dieye, personal communication, 2022). The database includes deal information such as deal type (agribusiness, mining, etc.), size, year of transaction/negotiation, or implementation status where available. While valuable, this database is not exhaustive and inherently contains a degree of uncertainty due to the dynamic nature of LSLAs, including new installations, abandonments, and expansions. As a result, it requires ongoing updates that are both time consuming and costly.

In this work, we focused specifically on three regions where 92% of ongoing land transactions and 100% of current and planned installations (Bourgoin et al., 2019) are concentrated (see **Figure 1.4**):

- 1- In the north, the arid Senegal River Valley is an important agricultural region with a growing number of LSAs, mainly focused on horticulture, sugarcane production, and cereals, mainly rice.
- 2- In the center, the sylvopastoral area of Ferlo, composed mainly of tree and shrub savannah, is home to most of the LSAs focused on gum arabic production.
- 3- Close to Dakar, the Niayes includes many LSAs dedicated to horticulture. The vegetation consists mainly of open agricultural parkland.

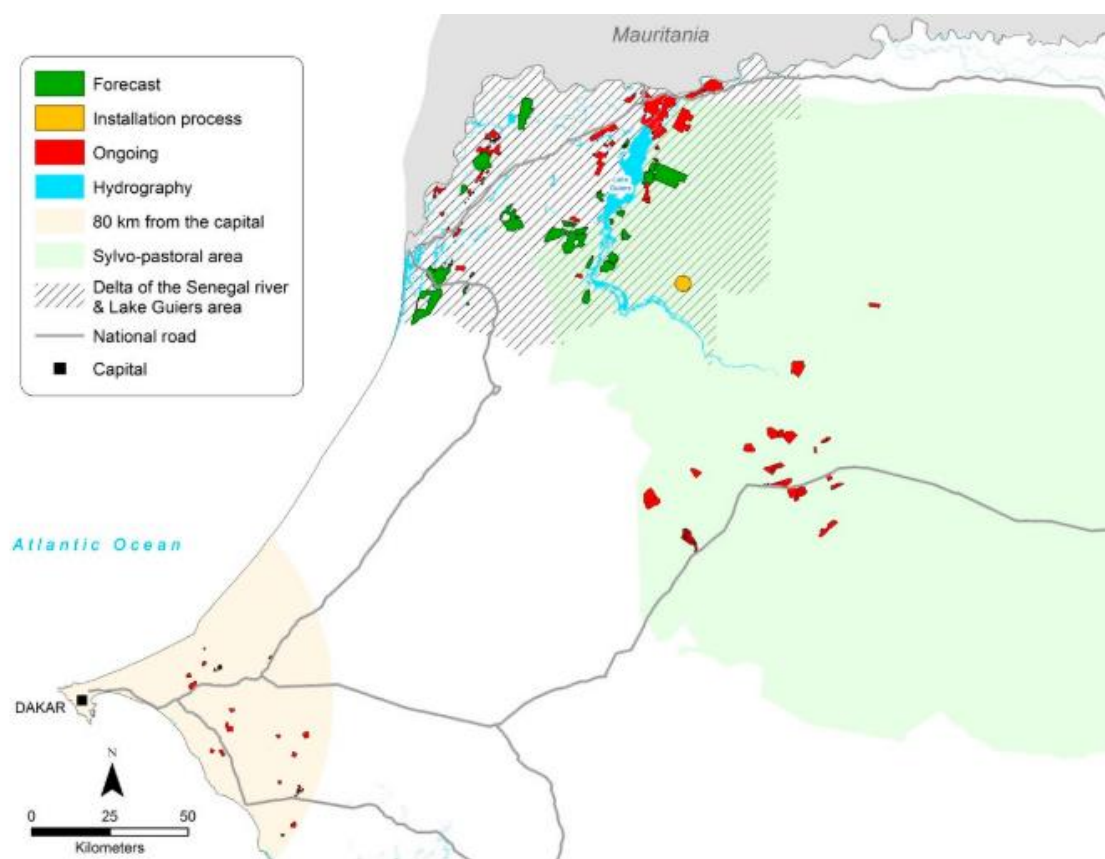


Figure 1.4: Distribution of on-going and intended agricultural LSLAs in the three main hotspots in Senegal : the Senegal River Valley (hatched area), the Ferlo (green area) and the extended area of the Niayes (pale orange). Source: (Bourgoin et al., 2019)

## 1.4.2 Research Objectives and conceptual framework

**This thesis aims to develop an unsupervised and generic framework designed to detect LULC changes at large scales potentially associated with Large Scale Agricultural Investments (LSAIs).**

As a result of the literature review conducted in Section 1.2, some common characteristics of LSAIs have been identified and are briefly described below. Newly implemented LSAIs are expected to: (i) induce changes in land surface vegetation composition, which may be reflected in vegetation index time series; (ii) rapidly occupy large areas, which may appear spatially as large hotspots of change; and (iii) practice intensive agriculture, which may be reflected in geometric shapes and a particular and ordered spatial arrangement of fields.

The strategy adopted (see **Figure 1.5**), based on these characteristics, consists of three main steps:

(1) the first one is the automatic detection and selection of the most significant pattern (i.e. mostly seasonal) change in long and dense VI time series that is most likely to be associated with anthropogenic LULC changes. **This is based on the hypothesis that most**

**anthropogenic changes induce a change in the time series spectro-temporal pattern.**

Change detection is first performed at the pixel level, using MODIS 250-m NDVI time series whose temporal (16 days) and spatial resolution (250 m) is optimal for LSAI detection.

(2) As the detected changes are not exclusively related to LSAI, the second step, still at the pixel level, aims to better understand the relationships between the main drivers of change and the time series types of change and to find a set of discriminative change metrics that can help to differentiate hotspots of LSAI-related changes.

(3) The final step is to automatically extract, classify, and characterize large hotspots of change through unsupervised classification to identify those potentially associated with LSAIs. This process relies mainly on evidence derived from the previously mentioned MODIS NDVI SITS derived spectro-temporal change metrics, associated with structural and textural features computed from higher spatial resolution satellite imagery (Landsat) for each change hotspot.

**The final product is a map of potential LSAIs and characterization.**

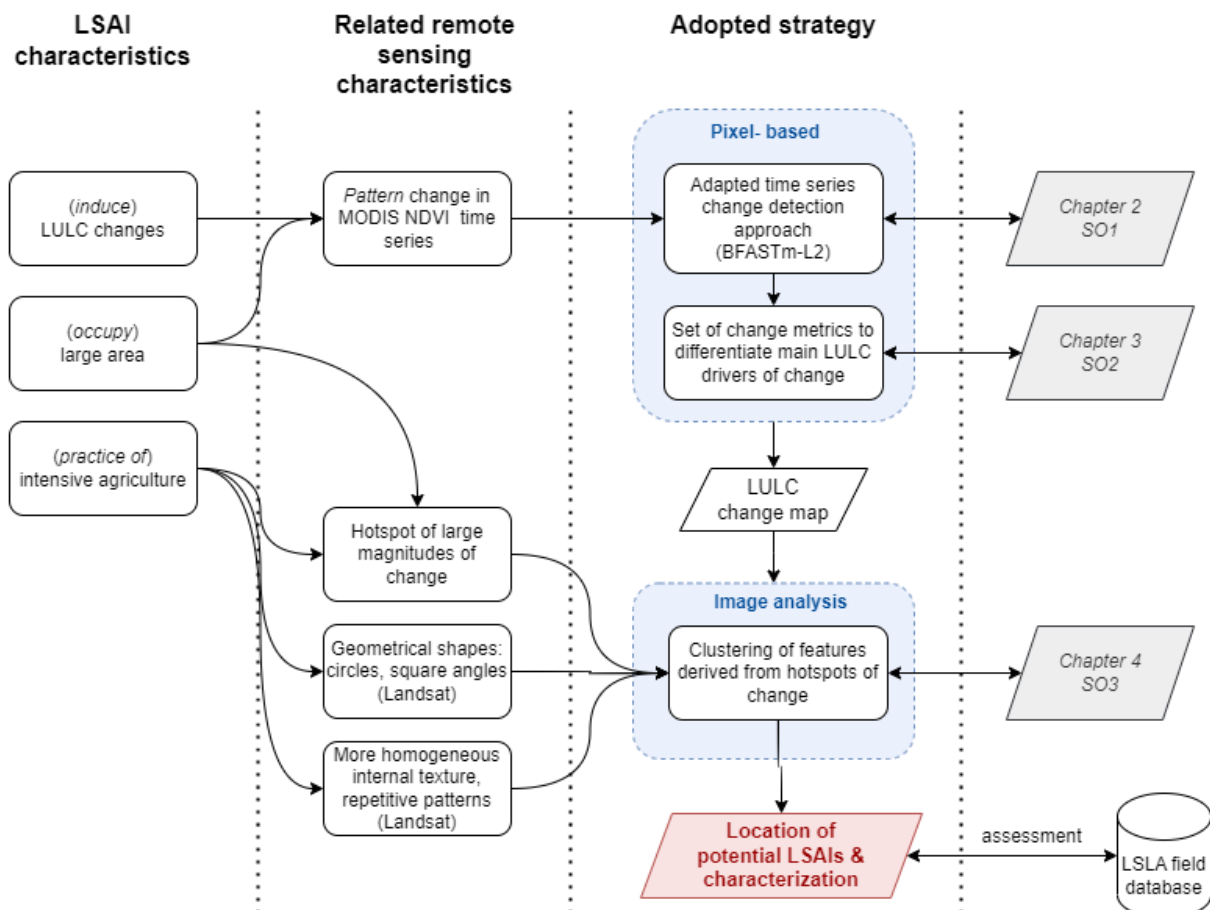


Figure 1.5: Thesis conceptual framework

## Chapter 1: Research objective

The main research objective has been subdivided into **three interrelated sub-objectives aimed at progressively building a body of evidence capable of refining the identification of LSAI-related LULC changes**. The sub-objectives are as follows:

Sub-objective 1: Automatic detection of potential anthropogenic changes using a change detection algorithm and medium resolution satellite (MODIS 250m) time series applied at a national scale

- RQ 1.1: Which method allows the detection and selection of the breakpoint causing the largest pattern (i.e. mostly seasonal) change in NDVI time series within long and dense time series?

Sub-objective 2: Gain insight and understanding of the changes most likely to be induced by climatic, natural, anthropogenic (non-agricultural) and agricultural LULC changes (including LSAIs)

- RQ 2.1: Are the main drivers of LULC change (i.e. climate variability, anthropogenic (non-agricultural), agricultural and natural changes) more likely to cause a particular type of change in NDVI time series?
- RQ 2.2: Which time-series derived change metrics may be useful to differentiate the different time series types of change?

Sub-objective 3: Identify and characterize potential LSAI-related hotspots of LULC change at different scales

- RQ 3.1: How can we automatically extract potential LSAI related hotspots of change?
- RQ 3.2: What spectro-temporal (from medium spatial resolution MODIS time series) and spatial features (from high spatial resolution Landsat imagery) are common to LSAIs and may distinguish them from other land dynamics?

### 1.4.3 Thesis outline

This thesis consists of six chapters, including this introductory chapter. **Chapter 2 (first accepted scientific paper) addresses the first research question (RQ 1.1)** by proposing a change detection approach, BFASTm-L2, which uses a high-speed algorithm (BFAST monitor) combined with a time-series similarity metric (Euclidean distance L2) sensitive to seasonal change to select the breakpoint associated with the largest seasonal change (hereafter referred to as magnitude of change) in long and dense SITS with multiple breakpoints. The proposed method was tested on two datasets (a published benchmark dataset composed of simulated SITS to test its sensitivity to different types of changes, and MODIS NDVI SITS over 200x200 pixel study sites with LSAIs in Senegal).

**Chapter 3 addresses RQ 2.1 and RQ 2.2** to gain insight and understanding of the LULC changes most likely to be driven by climatic, natural, anthropogenic (non-agricultural), and agricultural (including LSAs) changes. To do this, an RGB composite map was proposed that was created from three metrics of change whose sensitivity to different types of change was also tested on the aforementioned benchmark dataset: a time-series shape dissimilarity metric, an NDVI post/pre-change ratio, and the BFASTm-L2 magnitude of change. Map color (i.e., the signature of change), observed patch sizes and geometries, and ancillary data (including field information, precipitation data, NDVI time series analysis, and Google Earth imagery) were used to relate major drivers of LULC change to different types of VI time series change.

**Chapter 4 tackles RQs 3.1 and 3.2** through a three-step process applied independently to the Niayes and Senegal River (SR) regions, known for their high LSAI presence. First, indiscriminate LULC changes were detected at the national scale using the BFASTm-L2 algorithm applied to MODIS NDVI data. The magnitude of change was then weighted using the change metrics proposed in a previous work: the time series shape dissimilarity and the NDVI ratio, to highlight potential LSAs. Second, hotspots of change were extracted using data-driven contour analysis, and the main drivers of LULC change were identified using Google Earth imagery as well as field data on LSAs. Finally, for each of the extracted hotspots, spectro-temporal, and textural/structural features were computed from MODIS and Landsat data respectively and their effectiveness in discriminating LSAs was investigated through unsupervised analysis.

**Chapter 5 presents the main findings** of this thesis and reflects on the implications of the findings. Finally, **Chapter 6** concludes the thesis.

## 2 BFASTm-L2, AN UNSUPERVISED LULCC DETECTION BASED ON SEASONAL CHANGE DETECTION – AN APPLICATION TO LARGE-SCALE LAND ACQUISITIONS IN SENEGAL

---

This chapter is based on:

Ngadi Scarpetta, Y., Lebourgeois, V., Laques, A.-E., Dieye, M., Bourgoïn, J., Bégué, A., 2023. BFASTm-L2, an unsupervised LULCC detection based on seasonal change detection – An application to large-scale land acquisitions in Senegal. *International Journal of Applied Earth Observation and Geoinformation* 121, 103379.

## 2.1 HIGHLIGHTS

- Change maps based on breakpoint magnitudes often highlight abrupt/amplitude changes
- Land use changes induced by Large-Scale Land Acquisitions are mostly seasonal
- BFASTm-L2, a fast and unsupervised new method to timely detect seasonal changes
- BFASTm-L2 over performed BFASTm, BFAST Lite, Edyn mapping regional LSLAs
- BFASTm-L2 may be used to highlight important land use changes at regional scale

## 2.2 ABSTRACT

In the context of Global Change Research, detection, monitoring and characterization of land use/land cover (LULC) changes are of prime importance. The increasing availability of dense satellite image time series (SITS) has led to a shift in the change detection paradigm, with algorithms able to exploit the *full temporal information* laid down in *SITS*. So far, most of these algorithms have focused on the detection of abrupt and gradual changes, and thus developed breakpoint detection based on significant deviations from the mean. However, LULC changes may manifest themselves in other patterns, particularly changes in seasonality (amplitude, number and length of the growing seasons) that are harder to detect. In this paper, we propose a simple method to automatically select the breakpoint linked to the biggest seasonal change in long and dense SITS with multiple breakpoints. This approach - BFASTm-L2 - relies on linking a high-speed algorithm (BFAST monitor) with a time series similarity metric (Euclidian distance L2) sensitive to seasonal changes. The capacity of BFASTm-L2 to identify the date of change in different situations was tested on two data sets, and compared to the performances of three other algorithms (BFAST monitor, BFAST lite, and Edyn). The data sets are 1. a published benchmark data set composed of 25 200 simulated SITS with different change types and change magnitudes, and 2. the 2000-2020 MODIS NDVI SITS over a 200x200 pixels area in Senegal including different study sites which have undergone recent LULC changes due to agricultural large-scale land acquisitions (LSLAs) (as reported in the ground field database used in this study). The results show that BFASTm-L2 is efficient in accurately detecting in time most of the changes, and, in contrast with BFAST Lite and BFASTmonitor, to spatially highlight LSLAs-induced changes without the need of any prior knowledge. The automatic proposed approach, faster than BFAST Lite and Edyn, and with very few tuneable parameters, may thus be easily implemented in unsupervised pipelines to map and analyse generic LULC changes at regional scale.

## 2.3 INTRODUCTION

In the context of Global Change Research, detection, monitoring, and characterization of land use/land cover (LULC) changes are of prime importance. Global satellite-based Earth observation, with its repetitive coverage at short intervals and consistent image quality, has led to major advances in the field by providing insights into the land dynamics of large and remote areas. Recent studies have benefited from the increasing availability of free remote sensing data and the overwhelming increase in computing power to identify changes that occur over time. In particular, the availability of dense Satellite Image Time Series (SITS) has led to a shift in the change detection paradigm, with ever-increasing approaches and algorithms exploiting the *full temporal information* contained in *SITS*. Changes detected are usually categorized as “abrupt”, “gradual” and “seasonal”. While “abrupt” refers to short-term, large magnitude date-to-date changes (e.g., deforestation, fire, or urbanization), “gradual” (also referred as trends) refers to long-term (i.e. inter-annual), small magnitude date-to-date changes (e.g. land degradation, forest recovery) (Zhu, 2017). “Seasonal” changes are those affecting time series seasonality (i.e., vegetation phenology), and refer more explicitly to changes in the number of growing cycles (i.e., number of “peaks” of vegetation activity), in the season amplitude and in the length of season (i.e., growing season baseline length).

So far, change detection algorithms have mostly been used to detect significant deviations from the mean in forest ecosystems (Ochtyra et al., 2020) . This is explained by the fact that discrimination between phenological changes driven by climatic variability and disturbances is easier in stable seasonal environments such as evergreen temperate or tropical forests than, for example, in drier ecosystems (Browning et al., 2017; Gao et al., 2021; Zhu and Woodcock, 2014). Less focus has been given to other land surface dynamics such as changes in land use (Verburg et al., 2009). Land use, which refers to the purposes for which human exploit the land cover, is hardly inferred directly from remote sensing images and very often needs ground-knowledge to be accurately assessed. For this task, temporal series of Vegetation Indices (such as the Normalised Difference Vegetation Index or NDVI) are often used. While remote assessment of a single land use may be difficult to achieve, land use *changes* may however be assessed by detecting persistent seasonal changes (e.g. changes in the amplitude, length of season and/or number of seasons) within VI’s time series, as homogeneous land practices are expected to present typical intra-annual patterns (Setiawan and Yoshino, 2014).

This is how, Hentze et al. (2017), focusing on the detection of specific seasonal changes (from unimodal to bimodal distributions and vice-versa) in the NDVI, jointly with a seasonal-trend analysis, were able to identify agricultural land tenure transitions (from large-scale to small-



scale and vice-versa) in Zimbabwe using BFAST (Breaks For Additive Season and Trend; (Verbesselt et al., 2010a)). However, their method has the drawback of not being up-scalable because of the algorithm used (BFAST is computationally expensive), and because their need of external data (crop mask).

Speed, automaticity and accuracy of the detection of seasonal changes are thus crucial for the development of *generic* approaches to land use change detection. This is particularly relevant when using long and dense temporal series, in which a mix of different types of changes with different intensities may occur. As a consequence, approaches that perform temporal segmentation (such as DBEST (Jamali et al. (2015) ) or LandTrendr ( Kennedy et al. (2010))) which aim to find major abrupt and gradual changes without considering the seasonality are out of the scope of this study. The same applies to the deep learning methods, despite the growing interest in the remote sensing community due to their ability to automatically extract spectral-spatial features from satellite imagery (Tuia et al., 2021; Yuan et al., 2020; Zhu et al., 2017). Because 1- the unsupervised/ semi-supervised deep learning approaches for LULC change detection are still at early stage of development (Leenstra et al; Meshkini et al., 2021), in addition to the 2- poor interpretability of the models, and the 3- low generalization performance linked to the highly contextual-dependent methods, the labeled-data scarcity (hampering the method's scalability at regional scale), and the huge spectral-spatial variability of the targeted object (land use changes), these approaches are not considered in this study. They however should be the object of further research.

To continue with the change detection characteristics sought in this study, to be applicable on large scale, change detection approaches need to be ease to use (i.e number of parameters), and preferably integrated on cloud-computing platforms. Amongst the well-established statistical-based algorithms implemented on big data platforms, BFASTmonitor (Verbesselt et al., 2012) stands out from the rest because of its simple configuration, speed and massively-parallel GPU implementation (Gieseke et al., 2020), and its relative good performance in detecting seasonal changes (Awty-Carroll, 2019). Its major drawback lays in its high commission error (false positives). Minimization of false positives in long and dense time series is often performed using a threshold on the breakpoint magnitude (Gao et al., 2021), or by using some statistical tests such as the Chow test (Bullock et al., 2020). When more than one true breakpoint is found, selection of a unique breakpoint for mapping purpose may be difficult, unless when looking for specific changes (in magnitude, sign or pattern) in a specific period of time.

Selection of a unique breakpoint may be more difficult when using algorithms that do not perform a season-trend decomposition (e.g. BFAST monitor, EWMACD). These latter, usually

faster (and thus suitable for large-scale applications), are often able to detect any type of change, including the seasonal ones, within the same non-decomposed time series. However, because the computed breakpoint's magnitudes are linked to deviations from the model, selection of a unique breakpoint based on its magnitude may disfavour the selection of seasonal changes, particularly those modifying the time series "shape" (i.e. number of growing cycles per year, length of season) without heavily impacting its amplitude.

In this study, we propose a simple, fast, generic and unsupervised approach (hereafter referred to as BFASTm-L2) to select, in long and dense NDVI time series with multiple breakpoints, the optimal breakpoint linked to the most important land use change (i.e., linked to the most important seasonal or "pattern" change within the time series (in amplitude, length of season, NOS, or a mix of them)).

Two sets of research questions aimed to be answered in this paper: 1- The first one is broad and concerns the temporal accuracy (does the breakpoint correspond to the year of change?) and the sensitivity (how likely are different *type* of changes detected?) to different *types* and *intensities* of change of BFASTm-L2 and three change detection algorithms tested for comparison purposes: BFASTmonitor, BFAST Lite, Edyn. 2- The second set of questions concerns an application case, which is the detection (are the different change detection algorithms able to detect in long and dense time series, with likely multiple changes of different types and intensities, the land use changes related to LSLAs?) and mapping (which types of changes are most likely to be highlighted in maps using the breakpoint magnitude as mapping variable? Can LSLAs be pinpointed in such maps?) of Large-Scale Land Acquisitions (LSLAs) in Senegal.

More details on the material and methods used are given in the next section.

## 2.4 MATERIAL AND METHODS

### 2.4.1 Global approach

In this study BFASTm-L2 is proposed as a method to select, in long and dense NDVI time series with multiple breakpoints, the optimal breakpoint linked to the most important land use change, that is the one related to the most important "pattern" change within time series. Because the breakpoint selection is based on the breakpoint magnitude, we tested a magnitude metric based on the Euclidean distance (L2), a time series similarity metric proposed by Lhermitte et al. (2011), which is prone to be more sensitive to seasonal changes than to abrupt and gradual changes. BFASTm-L2 relies thus on the sequential running of

BFASTmonitor for the fast detection of breakpoints, jointly with the Euclidean distance (L2) for the breakpoint selection.

Because long and dense time series most of the time include multiple changes of different types (often combined) and intensities, a benchmark dataset of simulated seasonal time series including a unique change was used to answer the first research set of questions, aiming to assess the temporal accuracy and sensitivity of the tested change detection algorithms to different *types* and *intensities* of change. More specifically, a sub-sample of the benchmark dataset provided by Awty-Carroll et al. (2019) composed of 25 000 simulated SITS was used, including different single change types and multiple noise/change intensity levels (see Table 1). To answer the second set of questions, related to the performance of the change detection algorithms in detecting and mapping LSLAs in Senegal, the MODIS 16-day NDVI time series (MOD13Q1 v.6) acquired over the 2000-2021 period in different study areas with known LSLAs were used. Results were validated with ground observations. The flowchart of the data and methods is illustrated here below in **Figure 2.1**.

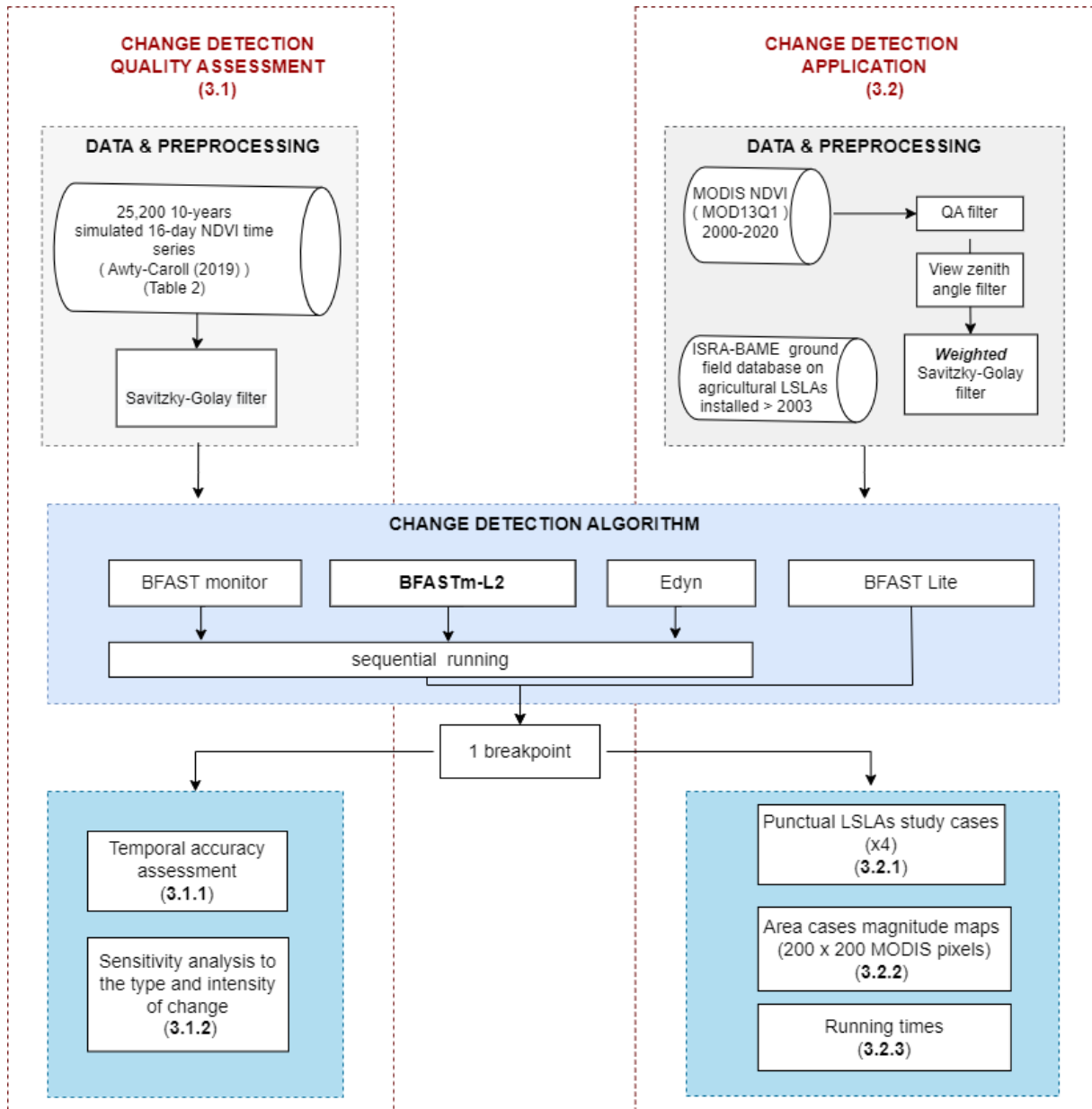


Figure 2.1: Flowchart of the data and methods

## 2.4.2 Data

### 2.4.2.1 Benchmark dataset and preprocessing

This study made use of the simulated NDVI time series (10-year 2006-2015 at 16-day temporal resolution) data set created by Awty-Carroll (2019) and available at <https://osf.io/taf9y/>. The 16-day temporal frequency makes this dataset suitable for the purposes of this study which uses the MODIS satellite imagery. In this study, only the gap-free simulations were used consisting of 25 200 seasonal time series generated to represent a large range of ecosystem dynamics based on the previous work of Verbesselt et al. (2010b). This dataset includes a large range of unique trend, abrupt (positive and negative), seasonal (amplitude, length of

season (LOS) and number of seasons (NOS)) changes, with eight levels of noise (random value from a normal distribution with a mean of 0 and a standard deviation  $\in [0: 0.01: 0.07]$ ), and 50 replicates (Table 2.1). The LOS changes are changes that move back the start of the season from 13 to 49 days. All changes were placed in January 2011.

Table 2.1. Types of change, levels of intensities and number of samples present in the simulated data set (Awty-Carroll et al., 2019). For the break/trend set, each abrupt change in NDVI is followed by either no trend or one of the six levels of trend present in the trend only set.

	Type of changes	Levels (units)	Number of simulations
-	No change	-	400
Trend	Trend only	$\pm [0.001, 0.0015, 0.002]$ (NDVI / year)	2 400
	Break ( <i>abrupt</i> ) / trend	$\pm [0.1, 0.2, 0.3]$ (NDVI)	16 800
Seasonal	Amplitude	$\pm [0.1, 0.2, 0.3]$ (NDVI)	2 400
	LOS ( <i>length of season</i> )	$- [13, 22, 30, 37, 43, 49]$ (days)	2 400
	NOS ( <i>number of season</i> )	1 to 2, 2 to 1	800
	<b>Total</b>	-	<b>25 200</b>

Because of the presence of noise hindering the detection of change, in this study the simulated time series were, as for the real MODIS time series smoothed using the Savitzky-Golay smoothing filter (Chen et al., 2004). A moving window length of 9 observations and a polynomial order of 3 were used.

#### 2.4.2.2 Study case dataset

##### 2.4.2.2.1 Study area and ground dataset

Senegal is in the west part of the Sudano-Sahelian zone and is characterized by an overall low average annual rainfall, but with high inter- and intra-annual variability that constrains the vegetation growth. The precipitation shows an increasing gradient along the North-South direction. A distinct seasonality is present, with a long dry season and a short rainy season spanning from late June to early October (Abel et al., 2019).

## Chapter 2: Material and Methods

As many African countries concerned by Large Scale Land Acquisitions (LSLAs) (<https://landmatrix.org>), Senegal had in 2016 around 3% of its total arable land (270 908 hectares) declared under contract by foreign investors from 12 countries (Harding et al., 2016). However, as more than 50% of those deals have an area under contract smaller than 5 000 ha, the country is less affected by the so called “megadeals” (>50 000 ha, only 2 out of 19 in 2016). Because of those climatic and LSLA characteristics, Senegal represents an interesting and difficult study case for the detection of land use system changes such as those induced by LSLAs.

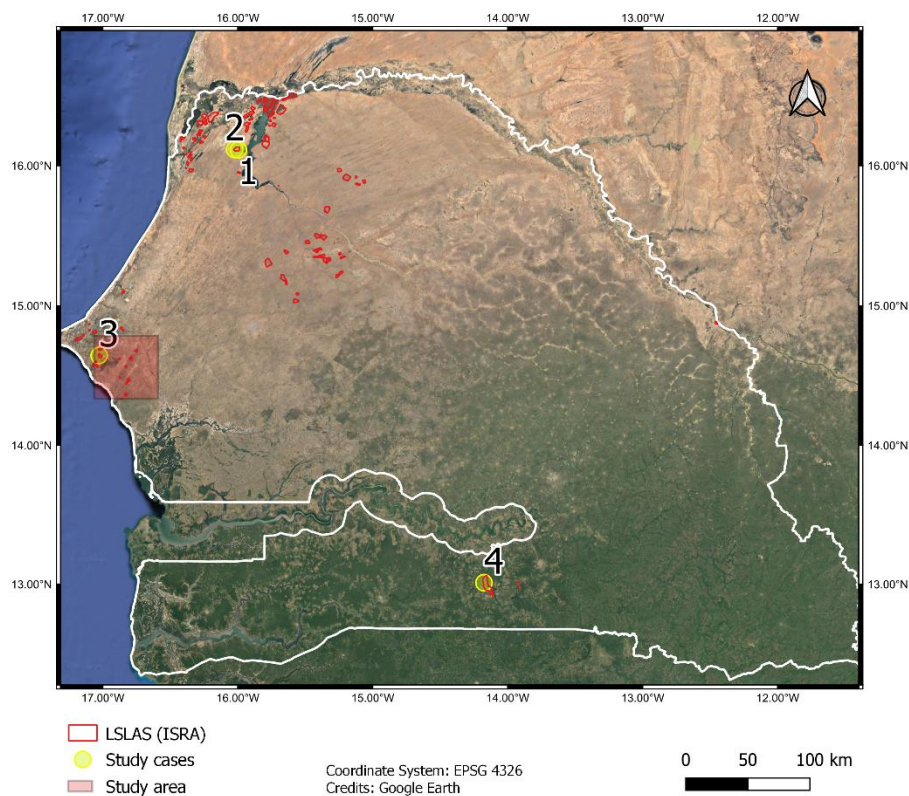


Figure 2.2. Study area (red box) and study cases (red dots) in Senegal located within agro-industrial concessions (red polygons) (source: ISRA-BAME field data base).

The Senegalese Institute of Agricultural Research (ISRA) conducted in 2019 an extensive field campaign on LSLA. More than 700 records and corresponding attributes were initially recorded in a database (M. Dieye, personal communication, 2022). Attributes consist in all kinds of information related to the identified deals, such as: name, deal type, coordinates, negotiation status, implementation status, year of transaction, size, previous land use, previous land tenure etc. From this database, and using observations from Google Earth/Sentinel Hub images, a selection was done to keep only active agricultural deals with growing crops (some deals are abandoned, others are still in negotiation), and with an implementation date after 2003 (year as from which changes may be detected from MODIS SITS, considering the training period length needed by the detection algorithms).

From this ground-field database subsample, four individual (points) LSLA study cases with different land processes under different ecoregions (see **Figure 2.2**) were selected for algorithm testing purposes. Point 1 is an example of a conversion from small agriculture to LSLA, points 2 to 4 are examples of conversions from natural vegetation to LSLA.

More specifically, study case 1 (16.1118°N, 15.9954°W) and 2 (16.1159°N, 16.0230°W) are in the Senegal river valley (rainfall: 150-600 mm; (Tappan et al., 2004)) and are within pivot irrigation areas belonging to a concession growing vegetables. Changes occurred from the end of 2011 for point 1, and the start of 2016 for point 2. Point 3 (14.6444°N, 17.0271°W), in the “Agricultural Expansion Region” as defined in Tappan et al. (2004) (rainfall: 600-700 mm), is within one of the production blocks belonging to a concession specialized in vegetable production, and changes are observed as from 2008. Finally, the southern point 4 (13.0104°N, 14.1766°W), in the Anambé basin (Casamance; rainfall: 800-1400 mm), is in the area of a concession specialized in irrigated rice. Field preparation is observed as from the start of 2007.

In addition to the LSLA study cases, an area of 200x200 MODIS pixels close to Dakar (red box in **Figure 2.2**), including six production blocks of different concessions (see red polygons in **Figure 2.7**), was selected for mapping purposes (see section 2.4 for more details). The area is spatially contrasted, with urban/rural areas and natural vegetation (see **Figure 2.7**).

#### 2.4.2.2.2 MODIS NDVI data and pre-processing

With its global coverage, moderate spatial resolution (250 m) and high temporal resolution (1 to 2 days), the Moderate Resolution Imaging Spectroradiometer (MODIS) sensor allows for the detection of subtle changes in land cover. A set of MODIS NDVI 16-day composites at 250 m resolution (MOD13Q1, collection 6) acquired over Senegal for the period 2000-2021 was pre-processed in Google Earth Engine. The 16-day composite NDVI product was chosen to reduce NDVI variability due to meteorological distortions like clouds. Series were gap-filled with linear interpolations and smoothed using an optimized weighted Savitzky-Golay filter (Chen et al., 2004) in order to reduce the noise. Weights were computed following the approach developed in Piou et al. (2013a), and are function of the pixel's reliability (i.e. quality flag, view zenith angle), and position in the moving window (exponentially decreasing weights with the distance to the window's centre). After some testing, a moving window length of 13 observations, and a polynomial order of 3 (in order to keep the ratio w/p close to 3-4) were used.



## 2.4.3 Methods

### 2.4.3.1 *Statistical-based change detection algorithms*

Up to now, the LULC change community has benefited from an ever-increasing emergence of change detection algorithms exploiting the *full temporal information* contained in SITS (Matthieu Molinier et al., 2021). Some of them are now well-established and implemented on big data platforms enabling fast processing of large volume data.

Change detection algorithms mainly differ in their approach to process time series and detect changes in them. In their approach to process time-series, two main groups of algorithms exist: While *offline* algorithms operate retrospectively on the complete time series, *online* (or real-time) algorithms aim to detect changes as soon as they occur (Bullock et al., 2020). Because offline methods make use of the entire time series, there are often more robust than online methods. In turn, online methods are often faster than offline methods as they only use some of the data preceding the real-time observation (i.e., training period). They are however prone to false positives and require a stable training period (limiting its use in places frequently disturbed (e.g. agriculture (Zhu et al., 2020)). As the length of the training period may have an impact on the quality of the fit - risk of overfitting when training periods are too short vs risk to include breakpoints in the training period when those are too long (Brooks et al., 2017))- , algorithms that are able to automatically select the optimal training period length (such as Edyn (Brooks et al., 2017), BFAST monitor (Verbesselt et al., 2012) or CCDC (Zhu and Woodcock, 2014)) are preferred.

In their approach to detect changes, while some algorithms provide the option of selecting an optimal model with one unique breakpoint (e.g. BFAST Lite (Masiliūnas et al., 2021)), almost all of them detect multiple breakpoints. Changes may be detected in the trend and seasonal component separately (e.g. BFAST (Verbesselt et al., 2010a), BEAST (Zhao et al., 2019)) or in the undecomposed time series (e.g. BFAST Lite, BFAST monitor , EWMACD (Brooks et al., 2014)), using temporal segmentation approaches based on residual-errors and angle criterion (e.g. DBEST ( Jamali et al. (2015)), LandTrendr ( Kennedy et al. (2010) )), or model-deviation seeking approaches (e.g. BFAST monitor, CCDC ( Zhu and Woodcock (2014)), EWMACD). Decomposition may be interesting for end-users with an a priori knowledge of the type of change foreseen. As an example, Mardian et al. (2021) assuming that pasture/rangeland conversions to cropland mostly impact VIs' seasonal component, applied a modified version of BFAST that constrained the model to detect only seasonal changes and obtained higher change detection accuracy compared to BFAST and BEAST. However, it is worth noting that algorithms that perform decomposition are usually slower than those that do not. In addition, errors in the decomposition (because of using an inappropriate model) may be translated in



errors in the accuracy of the changes detected in both SITS components (Mardian et al., 2021; Zhao et al., 2019). To avoid selection of a single-model algorithm, focus is recently given to ensemble learning algorithms such as BEAST (Zhao et al., 2019), that are however computational cost expensive. As one can see, selection of a single-model algorithm highly depends on the application scale, targeted type of change, environment, SITS source and frequency.

The approach proposed in this study for the selection of a single breakpoint linked to the biggest seasonal change in long and dense time series, is based on existing algorithms. For the selection of the change detection algorithm, focus was first given on speed, followed by the algorithm's sensitivity to seasonal changes and its ease of use (number of tuneable parameters). On these criteria, BFASTmonitor was chosen as the base algorithm for the change detection approach. For comparison purposes, in addition to BFASTmonitor, two other algorithms were selected: BFAST Lite, because of its robustness and speed (Masiliūnas et al., 2021), and Edyn, because of its speed and ability to capture seasonal changes (Awty-Carroll et al., 2019).

More details on the different change detection algorithms used and developed in this study are given in the next subsections and summarized in the Table 2.2 below.

## Chapter 2: Material and Methods

Table 2.2 : Main characteristics of the change-detection algorithms used in this study, their strengths and limitations

	Algorithm	Speed	Parameter number *	Model	Decomposition	Strengths	Limitations (others than speed and number of parameters)
Offline	BFAST Lite (Masiliūnas et al., 2021)	++	+	Model-deviation	None	Detection of abrupt and trend changes. Possibility to select a model with a unique (the highest-magnitude) breakpoint.	Low performance in capturing seasonal changes
	BFAST monitor (Verbesselt et al., 2012)	+++	+	Model-deviation	None	Sensitivity to seasonal changes (Awty-Carroll, 2019).	High false positive rate (Masiliūnas et al., 2021)
Online	EDYN (Brooks et al., 2017)	++	-	Model-deviation	None	Detection of subtle (sub-pixel) changes in long time series (model dynamically retrained); Able to capture seasonal changes	High sensitivity to algorithm parameters (Saxena et al., 2018)

\* Number of key parameters (others than (if applied) the order of the harmonic term, the probability threshold/ statistical significance level and the training period length):  
 <=2: +, >=3: -

## Chapter 2: Material and Methods

### 2.4.3.1.1 BFASTmonitor (BFASTm)

The BFASTmonitor (Breaks For Additive Season and Trend Monitor; (Verbesselt et al., 2012)) is an online (near real-time) unsupervised change detection algorithm that flags abnormal observations within a monitoring period, based on a stable history period. More specifically, once the start of the monitoring period has been defined, a stable history period is automatically selected using the reversed-ordered-cumulative sum (CUSUM) of residuals (default approach). Then, a regression model (here a linear harmonic regression model) is fitted based on the history period. Finally, the moving sums (MOSUM) of residuals are used (bandwidth defined by the  $h$  parameter) in the monitoring period to determine whether the model remains stable for new observations. A break is detected when the absolute value of the moving sums exceeds a probability threshold. The magnitude of change recorded represents the median of the difference between the data and the model prediction in the monitoring period. Because BFASTmonitor only needs a single observation to exceed the threshold, the algorithm is prone to false positives (Awty-Carroll, 2019; Ghaderpour and Vujadinovic, 2020) and magnitude thresholds are often applied to minimize them (Gao et al., 2021; Hamunyela, 2017). However, its massively-parallel GPU implementation (Gieseke et al., 2020), that makes it 14.5 times faster than the newly launched BFAST Lite (Masiliūnas et al., 2021), its implementation on Google Earth Engine (Hamunyela et al., 2020) and its good performance in detecting seasonal changes (Awty-Carroll, 2019), make of BFASTmonitor a potential change detection algorithm for the purposes of this study. This study made use of the Python package `bfast0.7`, available at <https://pypi.org/project/bfast/>, that provides a parallel implementation of the BFASTmonitor algorithm. The parameters used were:  $h=0.25$  (the MOSUM bandwidth),  $k=3$  (default and minimal number of harmonic terms possible), and threshold level=0.05. To detect multiple breakpoints, the algorithm was iteratively run on the MODIS NDVI 2000-2021 image stack, each 3 months, using a 3-year training period (after testing of two training period lengths  $h_1=2$  vs.  $h_1=3$  years) and a monitoring period of one year given that the BFAST monitor change magnitude is relative to the monitoring period length. No “penalty” period after each detected breakpoint (as opposed to Awty-Carroll (2019)) was applied to avoid missing significant changes. It is worth to note that this implementation of BFASTmonitor does not include the automatic determination of a stable training period yet.

### 2.4.3.1.2 BFAST Lite

The newly BFAST Lite unsupervised algorithm is built upon the BFAST algorithm (Verbesselt et al., 2010a) with the aim to improve its speed and flexibility (Masiliūnas et al., 2021). Compared to BFAST, this offline approach avoids the seasonal-trend decomposition and performs the model fitting in a single step, using a multivariate piecewise linear harmonic regression. In addition, it provides more robust statistics for breakpoint magnitude calculation,

such as the Root Mean Squared Deviation (RMSD) and the Mean Absolute Deviation (MAD), which are computed between the predicted values of the adjacent segments over the time span of one year before and after the detected break, as opposed to BFAST that computes the difference of the fitted value immediately before and after the break (Masiliūnas et al., 2021).

This study used the BFAST 1.6.1 R package available at <https://cran.r-project.org/web/packages/bfast/index.html>. Default parameters were used, with the number of harmonic terms equals to 3. The biggest breakpoint in magnitude was selected using the root mean squared deviation (RMSD).

### 2.4.3.1.3 Edyn: the dynamic version of EWMACD

EWMACD (Exponentially Weighted Moving Average Change Detection) is an online monitoring algorithm that aims to detect persistent subtle changes, such as forest degradation or thinning (Brooks et al., 2014). As BFASTmonitor, EWMACD uses a statistical control chart (here, the EWMA) on the residuals to detect deviations from the mean. In the original version, the algorithm does not retrain after having flagged a breakpoint. In Edyn, the dynamic version of EWMACD (Brooks et al., 2017), the harmonic model coefficients are dynamically updated, and the optimal training period length is automatically found based on the quality of the model fit.

Global parameters defined by the end-user are: 1- the parameter  $\lambda$  which defines the algorithm's robustness to low signal-to-noise ratios, and 2- the persistence that refers to the number of times flagged deviations must be successively detected for a change to be considered (in Edyn it is given as a proportion of a year). Finally, and with respect to the breakpoint magnitude, while originally given as a standardized value (i.e. the residual value divided by the chart's control limits, and then rounded down to the nearest integer value), the magnitudes in this study correspond to the residual values at the breakpoints' location. The reason for this choice is that, when ran continuously, more than one breakpoint may have the same standardized value impeding the selection of a unique breakpoint.

In this study the same values as in Awty-Carroll (2019) were applied ( $\lambda = 0.3$ , persistence=6). Their R version of Edyn (available at: <https://github.com/klh5/season-trend-comparison/tree/master/ewmacd>) was used, adapted from the original R script (Brooks et al., 2014) and Edyn (Brooks et al., 2017). This version enables selection of stable (without breakpoints) sliding windows of fixed length (2 years) that are used as training periods, thus allowing continuous monitoring. Compared to the original Edyn this adaptation does not find the optimal training period length based on the model fit quality.

### 2.4.3.2 BFASTm-L2: a new approach

The procedure to select the BFASTmonitor breakpoint linked to the biggest pattern change in long-term NDVI SITS is detailed in this section. The approach (hereafter BFASTm-L2) is based on the Euclidean distance (L2 distance) and is schematized in **Figure 2.3**.

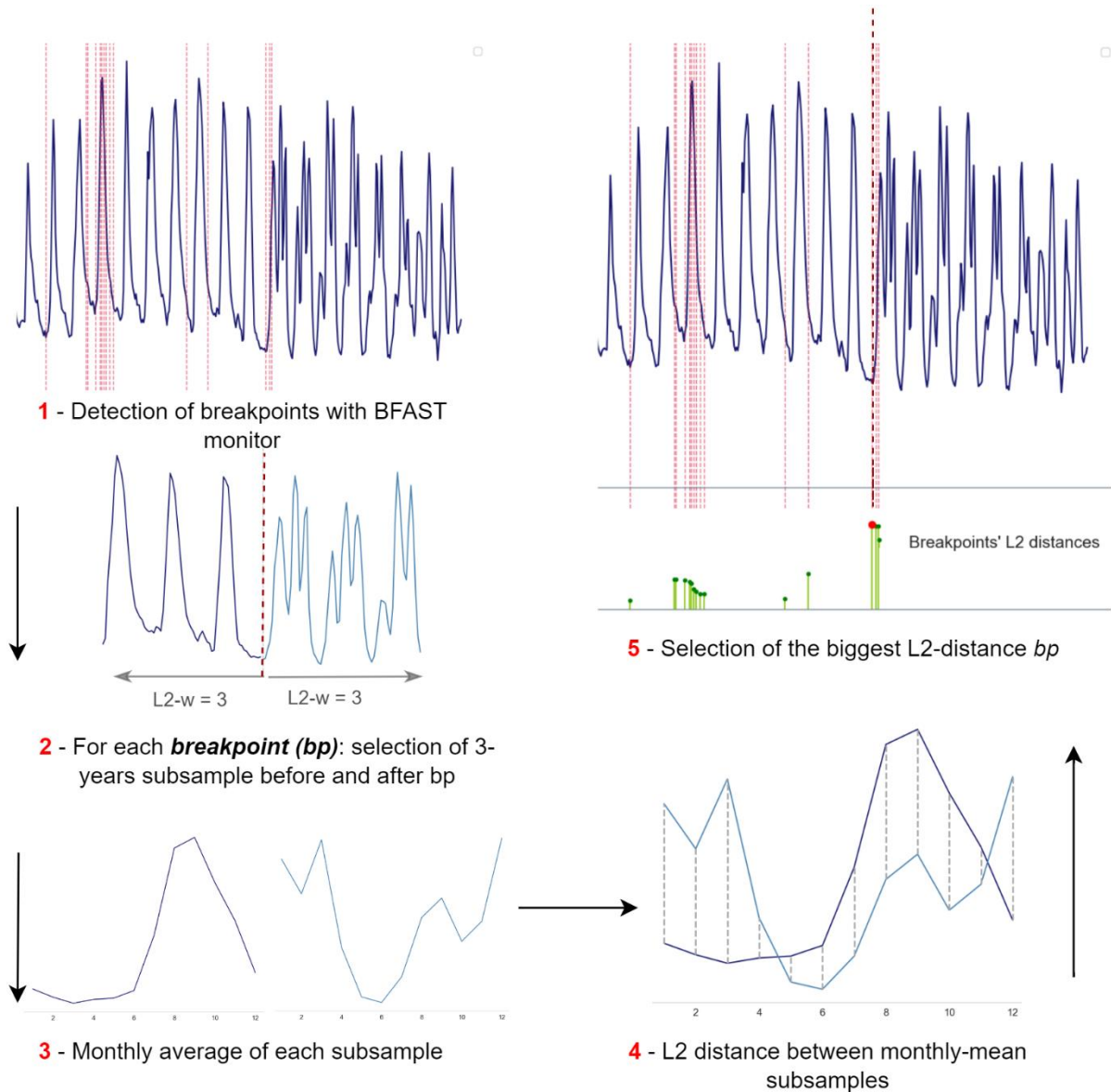


Figure 2.3. Flowchart of the BFASTm-L2 approach (in this example with  $L2-w=3$  years). Blue lines represent the NDVI time series subsamples (dark blue before breakpoint, and light blue after breakpoint). Detected BFASTmonitor breakpoints are presented with dashed red vertical lines.

In step 1 of **Figure 2.3**, BFASTmonitor is successively run each 3 months over the entire time series. Then, for each detected breakpoint, the time series segments  $L2-w$  ( $L2-w = 3$ ) before and after the breakpoint are extracted (step 2), and monthly averaged in annual subsamples (step 3). In step 4, the Euclidean distance between the 2 annual subsamples is computed using the Python `numpy.linalg.norm` function. Finally, the breakpoint with the highest L2 distance is selected (step 5).

This selection procedure is somehow similar to the method in Setiawan and Yoshino (2012), where a mean-based distance measure is computed between each two successive *annual* segments. It however differs in many aspects: 1- The distance metric is computed only where BFASTmonitor breakpoints are detected; 2- The segment length used here is 3 years (instead of 1 year), enabling to skip non-persistent changes mainly due to climate variability; 3- The (L2) distance used is more representative of an overall pattern change than the mean-based distance used in Setiawan and Yoshino (2012).

### **2.4.3.3 Evaluation of the algorithms' performance**

#### 2.4.3.3.1 Temporal accuracy to different types and intensities of change

The temporal accuracy of the single breakpoint detected by the four algorithms (BFASTmonitor, BFAST Lite, Edyn, and BFASTm-L2) was performed as following: 1- Each algorithm was run on each simulated time series, 2- the highest-magnitude breakpoint was selected (no selection performed for BFAST Lite), and 3- the absolute difference between the breakpoint date and the date of change (January of 2011) was recorded. Finally, bar charts of the selected breakpoints' date per type of change relative to the true date of change were plotted to assess the performance of each algorithm to accurately detect in time the different types of change.

#### 2.4.3.3.2 Breakpoint-magnitude sensitivity to different types and intensities of change

The breakpoint magnitude is often used to spatially detect significant hotspots of change. To unravel the type of change most likely detected by each algorithm (and thus spatially highlighted), the sensitivity of each algorithm to different types and intensities of change was assessed. For that, the distributions of the magnitude of the detected breakpoint were calculated per type of change, and represented using violin plots. Absolute values were used and normalized for each algorithm to allow comparisons between them. For all the change types (except the change in trend only, and the no-change category), only the breakpoints comprised between 2010 and 2012 (included) were used.

#### 2.4.3.3.3 Running times of the algorithms

The 5-run average times of BFASTmonitor, BFAST Lite, Edyn and BFASTm-L2, but also of the L2 distance metric alone computed sequentially each 3 months, were reported for areas of different sizes: 50x50, 100x100, 150x150, 200x200 pixels (the whole red box of **Figure 2.1**), using the MODIS NDVI 2000-2021 data set. Six CPU cores (parallel processing) were used on a 64 GB RAM computer.

### 2.4.3.3.4 Detection and mapping of LSLAs induced LULC changes: individual study cases

Evaluation of the change detection accuracy on the four real study cases presented in 2.1 was performed for: BFASTmonitor (training period of 3 years), BFAST Lite, Edyn (with  $\lambda = 0.3$ ) and BFASTm-L2 (using a training period length and L2-w size of 3 years), which were run continuously every 3 months. All the detected breakpoints, and selected highest-magnitude breakpoint were recorded (date and magnitudes).

Change maps based on the highest-magnitude breakpoint found in the 2000-2020 MODIS NDVI time series by BFASTm-L2, BFAST Lite, BFASTmonitor and Edyn were produced over the 200 x 200 MODIS pixels study area. The magnitudes for each map were then normalized between 0-1 in order to allow comparison. To quantitatively assess the performance of each method in detecting LSLAs induced changes, the average breakpoint magnitude within and outside the LSLAs (represented by red polygons in the study area) were computed and are presented in Table 2.3.

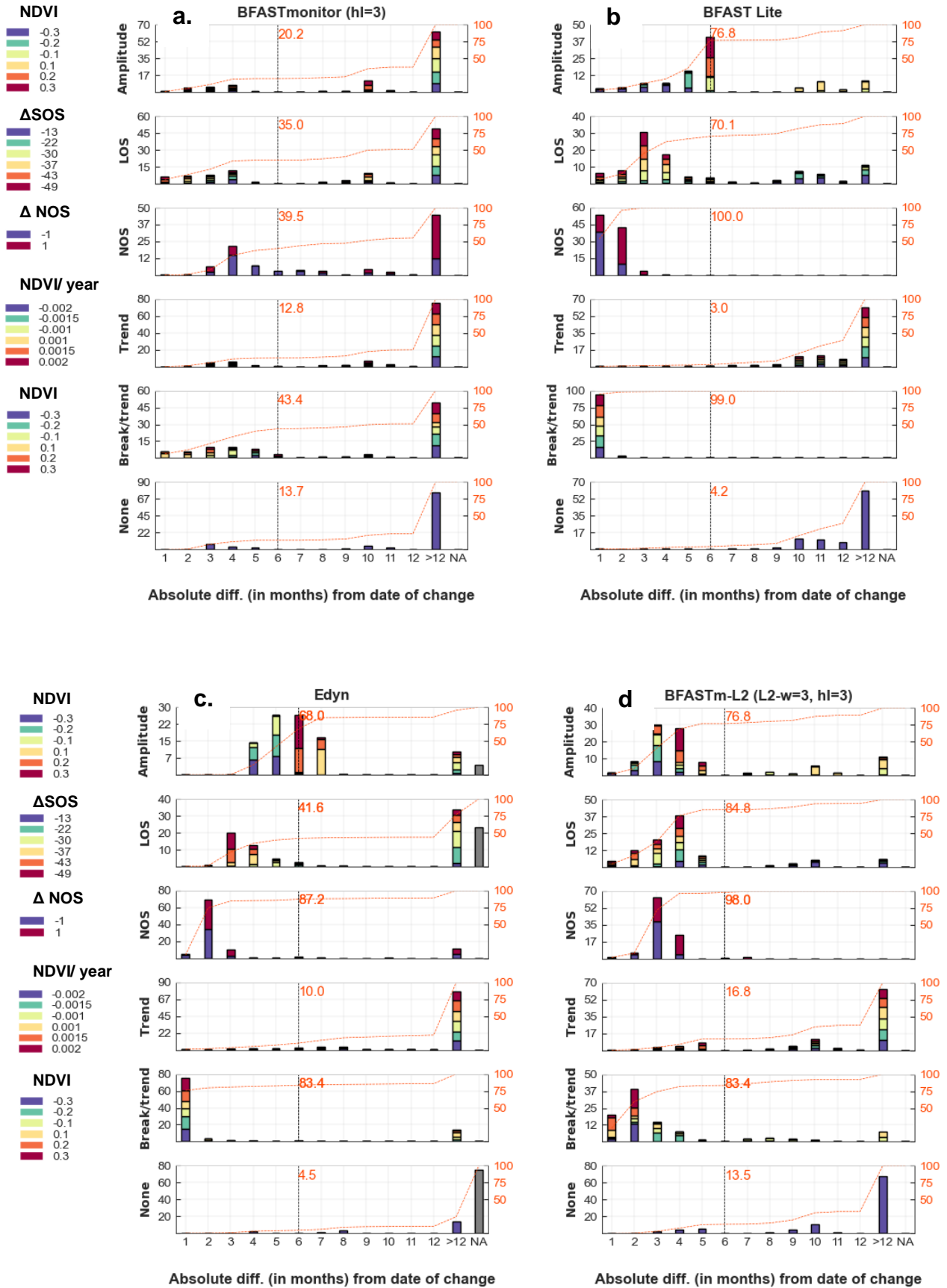
## 2.5 RESULTS

### 2.5.1 Temporal accuracy and magnitude sensitivity of the single breakpoint

#### 2.5.1.1 *Temporal accuracy of the breakpoint*

BFASTmonitor, BFAST Lite, Edyn and BFASTm-L2 were tested on the benchmark dataset (cf. section 2.1) to assess the performance of each method to accurately detect in time the different types of change (occurring in January of 2011). The breakpoints with the highest-magnitude were selected, and their temporal distributions relative to the true change are presented in **Figure 2.4** (cf. section 3.3.1).

## Chapter 2: Results





## Chapter 2: Results

*Figure 2.4. Temporal distributions of the highest-absolute magnitude breakpoint detected by BFASTmonitor (4.a), BFAST Lite (4.b), Edyn (4.c) and BFASTm-L2 (4.d), for six types of change (amplitude, length of season (LOS), number of season (NOS), monotonic trend (Trend), abrupt change (Break/trend) and no- change (None)). The x-axis represents the absolute difference in months between the breakpoint's date of change and the real date of change (January of 2011). The y-axis represents the proportion of samples detected at each x-axis unit. The stacked bars colours indicate the different intensities of change (see Table 2.1). The second y-axis (in orange) represents the cumulative percentage of the number of samples within each change type. A bar at 6 months was arbitrarily added to allow comparisons (% of samples with date of change  $\leq 6$  months from real date of change). The NA class represents the cases for which no breakpoint was detected.*

As a first observation one can see that most of the highest-magnitude breakpoints (of almost any change intensity), and with the exception of BFASTmonitor (4.a), are close in time to the true real change (less than 6 months). The algorithms have however slightly different sensitivity to the different types of changes:

- Regarding the seasonal changes (amplitude, LOS and NOS), BFASTm-L2 breakpoint magnitudes (4.d) are the most responsive to this type of change (>76% of the breakpoints in each subcategory of change are located  $\pm 6$  months of the true change). When looking more in detail, and compared to the other algorithms, BFASTm-L2 breakpoint magnitude is particularly sensitive to LOS changes (84.6% vs 70.1% for BFAST Lite and 41.6% for Edyn, at  $\pm 6$  months), and is as good as BFAST Lite in detecting NOS changes (98% vs. 100% at  $\pm 6$  months) and amplitude changes (76.8% for both at  $\pm 6$  months).
- In what concerns gradual changes (Trend), all the algorithms have at least one breakpoint induced by this type of change, mostly occurring one year before/after the true date of change.
- BFAST Lite (4.a) breakpoint magnitude is very responsive to the break/trend changes (99% of the changes detected within 6 months vs. 83.4% for both BFASTm-L2 and Edyn).
- Finally, regarding the no-change type, Edyn (4.c) was the only algorithm to correctly detect no change in almost 80% of the cases (as seen with the NA class).

Those first results show that selecting the BFASTm-L2 highest-magnitude breakpoint to correctly detect in time seasonal changes is effective. However, as one can expect multiple breakpoints detected in long and dense time series triggered by different types of change, the next question that comes up is: how comparable are the breakpoint magnitudes induced by the different change types?

As a small parenthesis, it is worth to note that the performances of Edyn, BFAST Lite and BFASTmonitor are different than in Awty-Carroll et al. (2019). This is explained by the fact that:

- 1- A subsample of the original dataset was used (dataset without missing data), and smoothed;
- 2- BFASTmonitor was run continuously without considering any "penalty" period after each detected breakpoint;
- 3- Only the distributions of the highest-magnitude breakpoints were evaluated;
- 4- The "correct" detection period considered here for the abrupt changes is larger

than in Awty-Carroll et al. (2019), who considered a maximum period of 3 months for the abrupt changes and 1 year for the seasonal ones.

2.5.1.2 Breakpoint-magnitude sensitivity to the type and intensity of change

**Figure 2.5** shows the violin plots of the breakpoint’s magnitude for each algorithm and type of change, as explained in 2.4.2. This type of representation helps in the identification of the type of change that will most likely be highlighted on a map when using the breakpoint magnitude as the mapping variable.

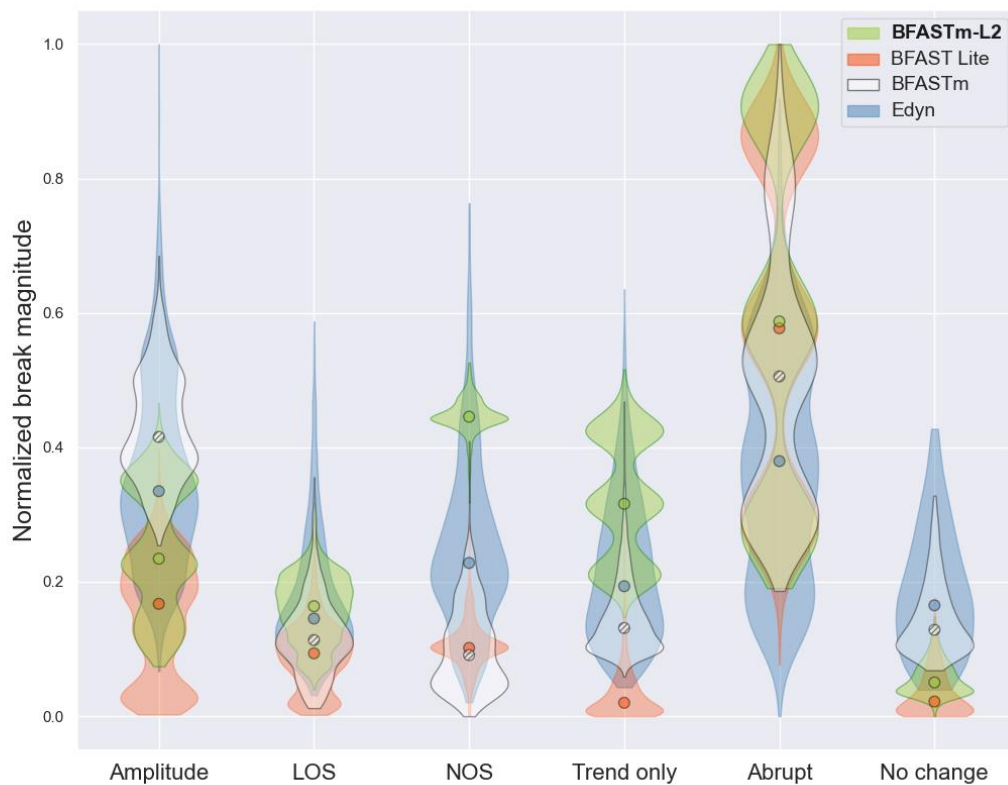


Figure 2.5. Violin plots of the normalized highest-magnitude breakpoint, per type of change for: BFASTm-L2 (green), BFAST Lite (orange), BFASTmonitor (white) and Edyn (blue). Dots represent the mean of each distribution. “Abrupt” refers to the Trend/break data subset without any trend change (refer to section 2.2.1).

Some general findings can be drawn from **Figure 2.5**. First, one can observe that all the algorithms, with the exception of Edyn, have their highest breakpoint magnitudes (>0.4) induced by medium to large abrupt changes (i.e. the trend/break data subset without any associated trend change) (>0.2 NDVI units; cf. Table 2.1). Detection of seasonal changes will therefore very likely be hindered in time series with large abrupt changes. Second, and with the exception of the no-change category, the lowest magnitudes are globally associated to the LOS changes. As such, one can expect LOS changes to be hardly detected in breakpoint-magnitude change maps.

Going more in depth with the algorithms' analysis, one can see that the BFASTm-L2 distributions means are the highest for two out of three (LOS and NOS) of the seasonal changes, with a remarkably high average for the NOS changes. In addition, the LOS and NOS distributions are well separated from the "No-change" category. This is important as it ensures the ability of the algorithm to highlight seasonal changes in breakpoint magnitude-based change maps. As a reminder, this remains possible as long as there are no large abrupt changes, or moderate to larger gradual changes (trends). Indeed BFASTm-L2, contrarily to BFASTmonitor or BFAST Lite, is sensitive to trends. Regarding the seasonal changes, BFASTmonitor, BFAST Lite and Edyn respond much better (higher magnitudes) to the amplitude changes than to NOS/LOS changes. This is particularly true for Edyn and even more for BFASTmonitor, which distribution base (for the smallest change intensity) is at almost 0.3.

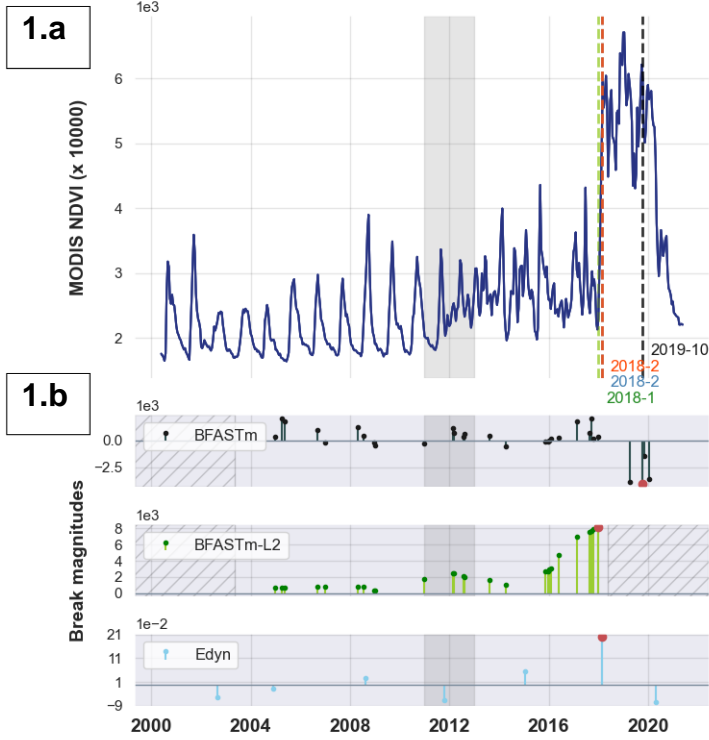
Finally, it is important to remember that the values represented in **Figure 2.5** were obtained from short and smoothed simulated time series, containing a unique change. As a result, the "No change" values for the different algorithms could in fact be higher in real time series with multiple small types of change, thus impacting the performance of each algorithm. To assess this, the different algorithms were tested on real time series and over a small area. Results are presented in the next section.

### **2.5.2 Detection of LSLAs LULC driven changes**

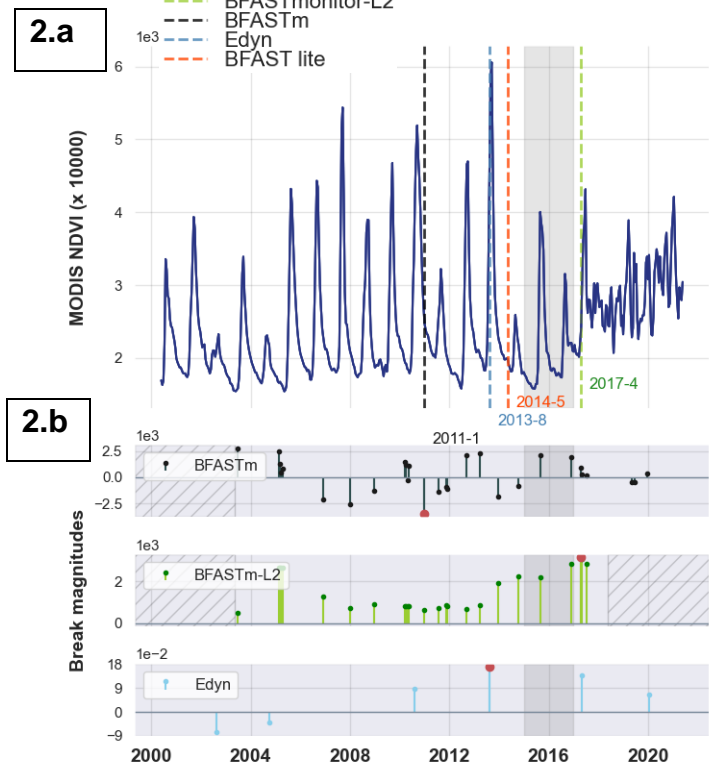
#### **2.5.2.1 Assessment on individual (pixel) study cases**

In this section, the capacity of the four algorithms to detect real seasonal changes on the four study cases (distributed from arid to humid conditions, see **Figure 2.1**) is evaluated. The overview of **Figure 2.6**, on which are overlaid the 2000-2021 MODIS NDVI time series and the period of change (grey boxes) for each study case, shows the diversity of type and intensity of changes found in the dataset: seasonal changes (all study cases present changes in amplitude, and study cases 1 to 3 show important NOS changes), an extreme abrupt change ( $\sim +0.25$  NDVI units) for study case 1 in 2018, and a strong positive trend (study case 2, start in 2016).

Study case 1



Study case 2





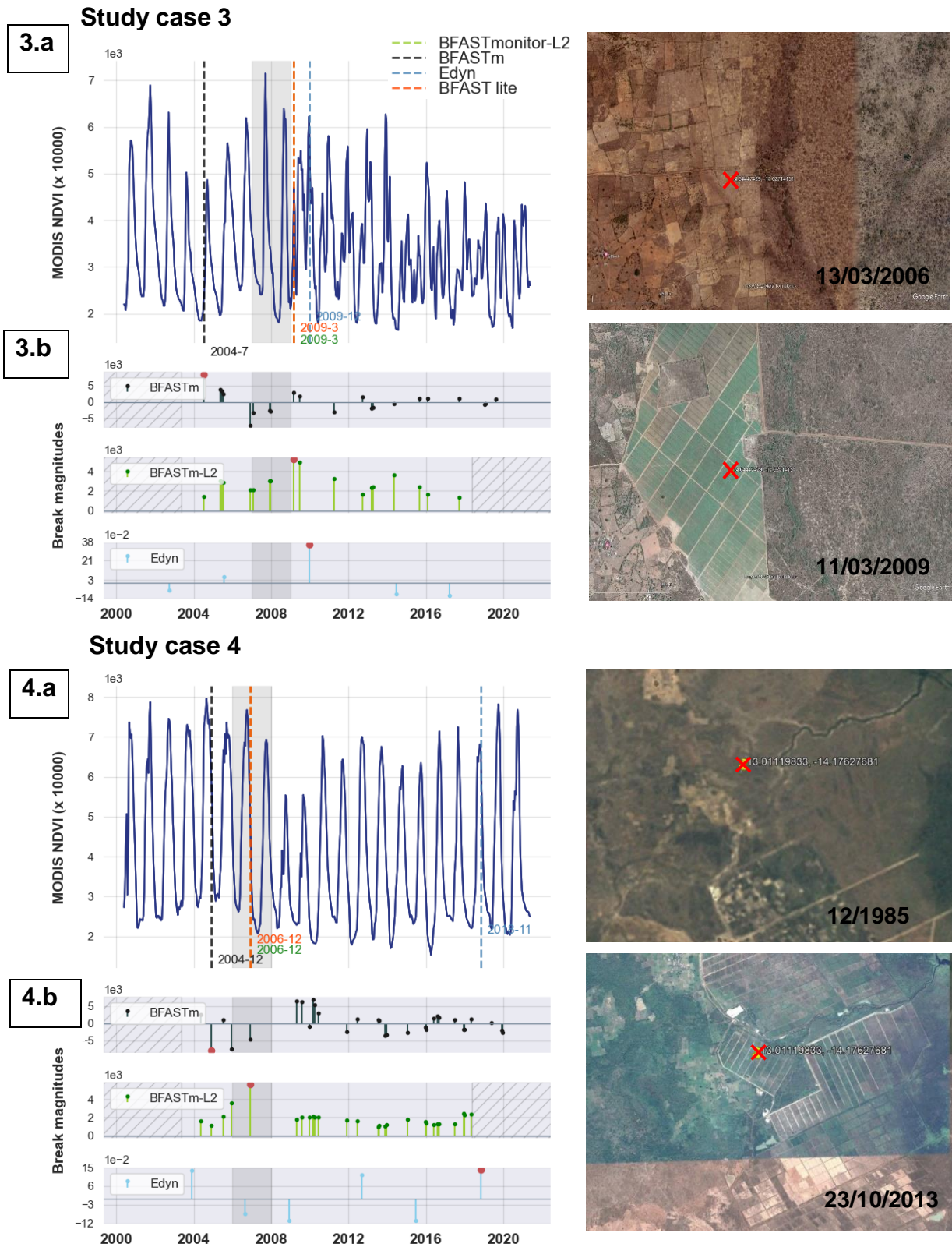


Figure 2.6. Application of the change detection algorithms to four Senegalese study cases showing agricultural LSLA implementations. For each study case, the corresponding pre-processed MODIS 2000-2021 time series with the different breakpoints detected (BFAS Lite)/selected (all others) are presented in the a subparts of the figures at the left. Figures at the right correspond to Google Earth snapshots closest in time with the LSLA implementation: the top/bottom snapshot for to the closest date available before/after the change. In the left figures: dashed lines correspond to the highest-magnitude breakpoints: green for BFAS Tm-L2, orange for BFAS Lite (1 break), black for BFAS Tmonitor and blue for Edyn (break dates are also given in the same colors to facilitate identification when

## Chapter 2: Results

*breakpoints are superimposed). Grey shaded areas correspond to the period of change (1 year before and after the observed date of change). Grey subplots (subparts b) at the bottom present all the breakpoints detected by BFASTmonitor (black), BFASTm-L2 (green) and Edyn (blue), along with their magnitude (height of vertical lines). Red dots pinpoint the highest-magnitude breakpoint (in absolute value). Hatched areas correspond to periods without monitoring (because of the initial period of time needed for training, and for computation of L2).*

On **Figure 2.6** (subpart b of each subplot), the multiple breakpoints detected by BFASTmonitor, BFASTm-L2 and Edyn are represented with vertical lines. The selected breakpoints are the ones having the highest magnitude, and are highlighted with a red dot at the extremity. BFAST Lite, BFASTm-L2 and Edyn performances (position of the unique breakpoint) are overall similar, with the following exceptions: i) a gap of 3/4 years is observed between BFASTm-L2 and BFAST Lite/Edyn in study case 2 and, ii) Edyn selected breakpoint is far away (>10 years) BFASTm-L2/BFAST Lite breakpoints in study case 4. Compared to BFASTmonitor, Edyn found significant fewer breakpoints, mainly due to the used-fixed period length needed after each detected breakpoint to ensure stability for model retraining.

In study case 1, two major events occur: a central-pivot irrigation system was installed at the end of 2011 (conversion from small to large-scale agriculture), and as from 2018 a huge increase in productivity is observed. The first change mostly translates as seasonal changes (NOS, amplitude), while the second induces a sharp abrupt change in 2018. In this case, BFAST Lite, BFASTm-L2, but also Edyn, output a higher-magnitude breakpoint for the abrupt change than the seasonal one, which is in line with the findings made in section 4.1.2 (**Figure 2.5**).

Study case 2 (conversion from natural vegetation), is also concerned by the installation of a central-pivot irrigation system at the beginning of 2016. It translates in the signal as seasonal changes (NOS, amplitude), with a positive trend. In this case BFAST Lite and Edyn selected breakpoints are closely located in time (~1 year of difference), and both differs (more than 3 years) from BFASTm-L2 selected breakpoint. While BFAST Lite and Edyn breakpoints are closely located from big changes in amplitude, BFASTm-L2 highest magnitude breakpoint is located at the beginning of 2017, when seasonal changes are most marked, and most probably related to the concession implementation and start of activities.

Study case 3 is another example of transition from natural vegetation to intensive agriculture, but in a less arid environment. The transition is visually observed in 2008 and mostly related to seasonal changes (NOS). Despite the high variability exhibited in the time series (particularly after 2008), BFAST Lite and BFASTm-L2 detects the same breakpoint, while Edyn differs from 9 months away in the middle of the second growing cycle of the first year of production.

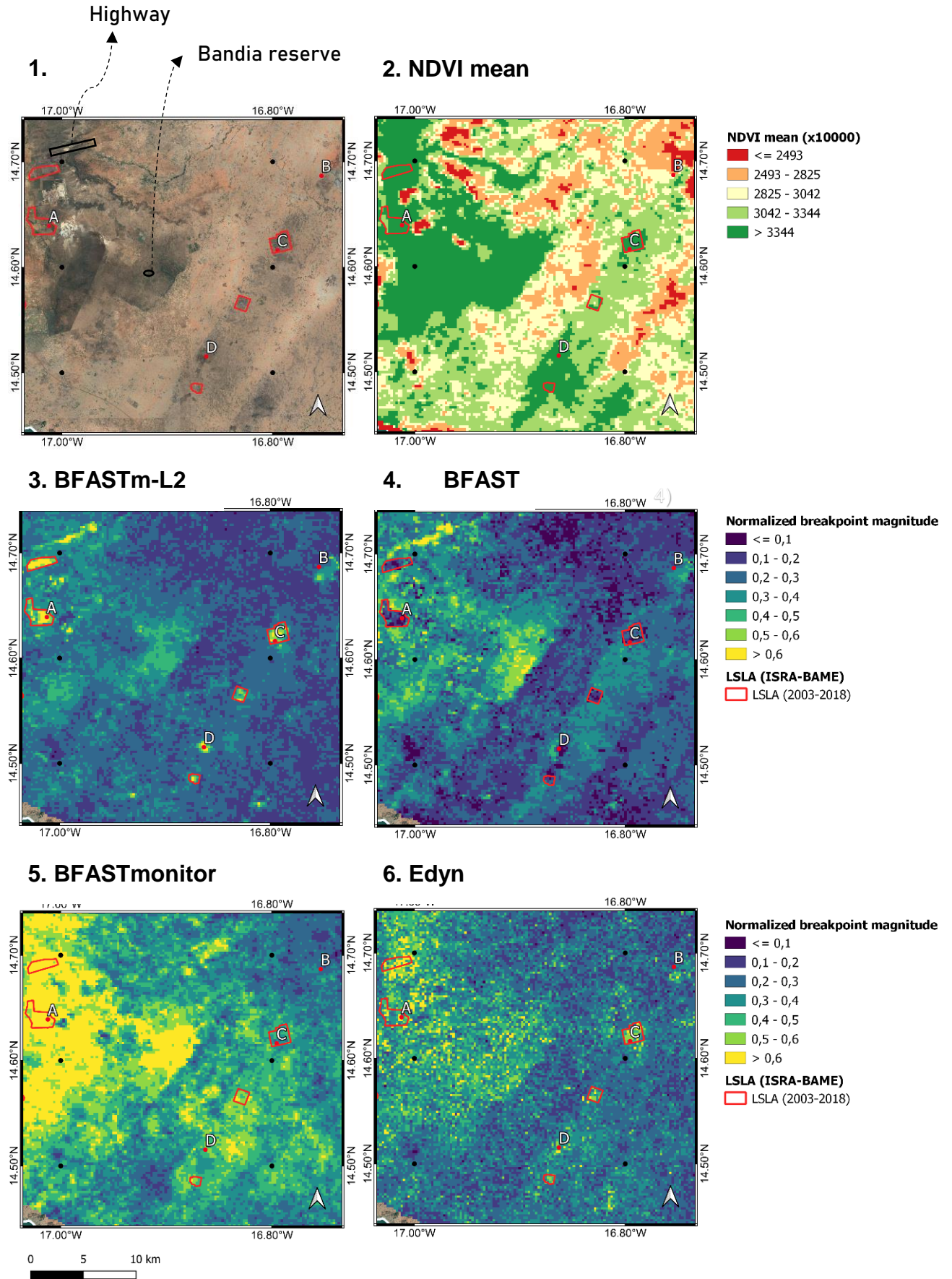
Lastly, study case 4 located in the most humid environment, shows more stability than the other study cases. Indeed, the land use change does not induce a significant change in the time series shape. The change being sought translates in a significant change in amplitude

and a negative trend. Here again, BFAST Lite and BFASTm-L2 agreed in the breakpoint detected, inside the sought period of change. Edyn diverges from many years away, and outputs its highest-magnitude breakpoint where an amplitude change is observed. As a reminder, the Edyn breakpoint magnitude used here is the residual of the detected breakpoint. While the standardized magnitude should be more adapted for the comparison of breakpoint magnitudes, it was not used here as more than one breakpoint may share the same maximum value thus hindering the selection of a unique breakpoint. Edyn sensitivity to the parameter lambda was tested, with the parameter set to 0.6 (Appendix A). All the highest-magnitude breakpoints found differed from those presented in **Figure 2.6**, including for study case 1 which contains a huge abrupt change.

### ***2.5.2.2 Detection of LSLAs induced LULC changes in breakpoint magnitude maps***

So far, it has been shown that the performances of BFAST Lite and BFASTm-L2 in detecting in time seasonal changes are similar. However, because of the different sensitivities of the breakpoint magnitudes to the different types of change ( **Figure 2.5** ), differences in the change maps based on the breakpoint magnitudes are expected. The change maps obtained with the different algorithms were evaluated over a study area close to Dakar, in which several LSLAs have been implemented since 2003 (**Figure 2.7**).

**PART I**





## PART II

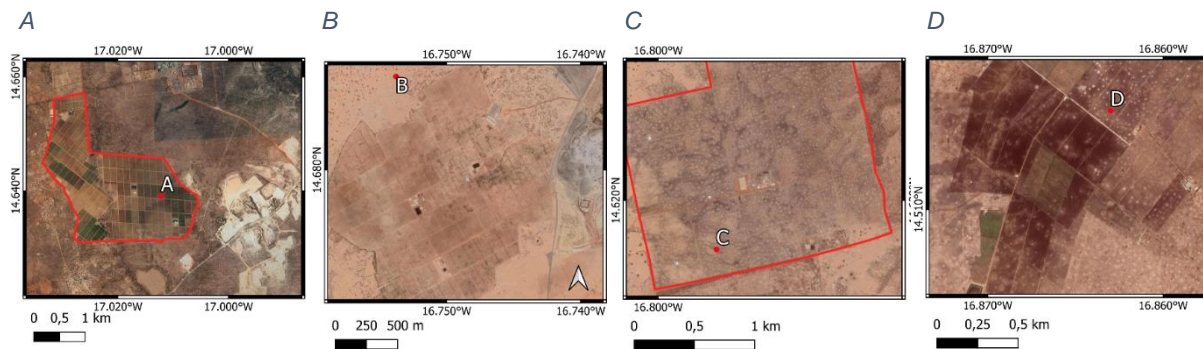


Figure 2.7. Change detection maps for each of the four change detection algorithms. .PART I) From left to right, top to bottom: 1) Snapshot view of the study area (Map data ©2015 Google), (2) MODIS NDVI average over 2000-2020, 3-4-5-6) Breakpoints (normalized) magnitude maps using: (3) BFASTm-L2, (4) BFAST Lite, (5) BFASTmonitor, (6) Edyn. Red polygons represent the active agricultural LSLAs reported in the ISRA ground-field database. PART II) Google Earth zoom-in of points A, B, C, D plotted in all the maps of PART I.

The change maps based on the breakpoint magnitude display different spatial patterns. While LSLAs (red polygons) in the BFASTm-L2 (subplot I.3) change map are clearly highlighted and have significantly higher magnitude values than the background (highest difference for BFASTm-L2 as shown in Table 2.3), this is not the case for BFAST Lite (subplot I.4) which difference between the values inside and outside LSLAs is null.

BFAST Lite (subplot I.4) highest magnitudes are mainly located in two areas of the map: 1- in the centre of the map, in the wooded vegetation of the reserve of Bandia (see Appendix B.1) and, 2- in the top left of the map, forming a stripe pattern of high values related to a highway construction in 2016 (see Appendix B.2). When looking more in detail the time series of these two places, the first change corresponds to an amplitude change (what seem to probably be the end of a recovery phase), while the second correspond to an abrupt change. This is in line with the results shown in **Figure 2.5**, which indicate that BFAST Lite breakpoint magnitude is mostly sensitive to abrupt changes and in a lesser extent to amplitude changes. Other seasonal changes (NOS and LOS) are not expected to be well discriminated from the background, particularly if, as it can be observed in I.4, the real background values are higher than what was expected from the no-change category in the simulated data set (<0.1 vs. 0.1-0.3).

At the opposite, and as mentioned above, BFASTm-L2 change map (I.3) based on the highest-breakpoint magnitude is able to highlight reported LSLAs such as those represented by polygons A (corresponding to the study case 3) and C (also see zoom-ins II.A and II.C), but also to detect newer LULC changes located in points B and D, and induced by agricultural activities as observed in the subplots II.B and II.D of **Figure 2.7**. “Pattern” changes induced by

## Chapter 2: Results

activities other than agriculture are also detected, such as the ones related to the highway construction at the north-west of the map.

Table 2.3: Average breakpoint magnitude (mean  $\pm$  standard deviation) in LSLAs (red polygons in the study area) and the whole study area

	BFAST Lite	BFASTmonitor	BFASTm-L2	Edyn
Breakpoint magnitude average in LSLAs (mean $\pm$ std)	0.22 $\pm$ 0.14 (n=258)	0.60 $\pm$ 0.15 (n=262)	0.50 $\pm$ 0.13 (n=262)	0.46 $\pm$ 0.16 (n=258)
Breakpoint magnitude average outside LSLAs (mean $\pm$ std)	0.22 $\pm$ 0.10 (n=37,653)	0.38 $\pm$ 0.13 (n=37,343)	0.23 $\pm$ 0.07 (n=37,343)	0.28 $\pm$ 0.10 (n=37,684)
Difference (in-out) (mean $\pm$ std)	0.0 $\pm$ 0.17	0.22 $\pm$ 0.20	<b>0.27 <math>\pm</math> 0.15</b>	0.18 $\pm$ 0.19

Regarding BFASTmonitor change map (I.5), one can see that the highest-breakpoint magnitudes are globally positively correlated to average NDVI (I.2). Even if the average of the magnitudes within LSLAs is high (0.6, the highest of the set; cf. Table 2.3), the magnitudes are not particularly sensitive to the changes induced by LSLAs. As an example, and in contrast with the BFASTm-L2 change map, magnitudes are overall low in points C and B. From **Figure 2.5**, and considering that the correlation with NDVI is lower for the other algorithms, we can advance that the BFASTmonitor magnitudes are mostly responsive to amplitude changes (instead of abrupt changes which would also be highlighted by BFAST Lite), or as observed in the previous section, to trend changes.

Finally, one can see in subplot I.6 that LSLAs are somehow discriminated in the Edyn residual-based change map (I.6). This is corroborated by the average difference of the magnitudes within and outside LSLAs, which even if lower than the BFASTm-L2 one, is still non-negligible (0.18 vs. 0.27, cf. Table 2.3). Despite the fact that the breakpoint residuals are not adapted for the breakpoint selection approach applied in this study (as the standardized magnitudes would be), the algorithm proved to respond to seasonal changes (in particular to amplitude changes). The drawback of the algorithm, and in particular for large-scale applications, is however its high sensitivity to the lambda parameter (Saxena et al., 2018) and low speed. This last point will certainly be improved in the future through the implementation of the algorithm on cloud platforms.

### 2.5.2.3 Running times of the algorithms

To assess the algorithm speed on real time series, the four algorithms and the L2 distance alone were applied on varying-size areas: from 10x10 to 200x200 pixels. Results are presented in **Figure 2.8**.

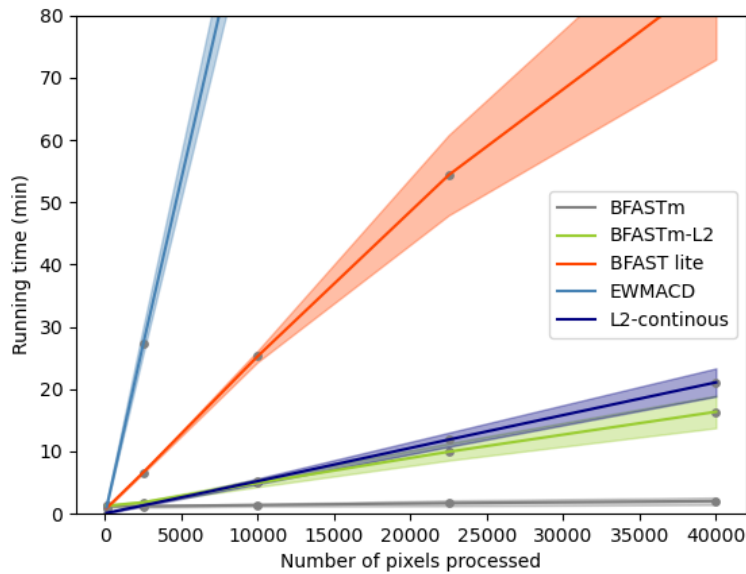


Figure 2.8. Running time mean and standard-deviations (5 runs) over areas with varying-size, using 6-cores parallel processing (64 Go RAM) for the 5 change detection approaches.

From **Figure 2.8**, BFASTm-L2 appears to be very fast (16.4 min for a 200x200 pixels area), just after BFASTmonitor (2.0 min for a 200x200 pixels area, with a median of numbers of breakpoints of 37). The “force-brute” method (the continuous computation of L2 at a 3-month step) took 21.1 min for the same area size, followed by BFAST Lite (86.5 min), and finally Edyn (439.8 min). Worth is to note that the R implementation of Edyn was called from python using the rpy2 interface, which may slower the entire process. Faster python implementation of Edyn may be further tested (as the very recent pyEWMACD available in github <https://github.com/lewistrotter/pyEWMACD>).

## 2.6 DISCUSSION

### 2.6.1 BFASTm-L2, an efficient method to detect seasonal changes

The results obtained with the single-change simulated data set showed that BFASTm-L2 was the approach with the overall best performance in accurately detecting seasonal changes, thereby demonstrating the efficiency of combining BFASTmonitor and L2 distance for timely breakpoint selection (**Figure 2.4**). The algorithm was particularly good at detecting, in a benchmark dataset with single changes, LOS changes (84.8%) and NOS changes (98%), but

also changes in amplitude (76.8%) and break/trend changes (83.4%), which were however better detected by BFAST Lite (99%).

When tested over simulated time series with a unique change, all the algorithms showed increased sensitivity (i.e. higher breakpoint magnitude) to abrupt changes (**Figure 2.5**). However, when considering the seasonal distributions alone in the violin plots of **Figure 2.5** BFASTm-L2 presented the highest magnitude means for LOS and NOS. This was also a goal of this study, as it enhances the probability of being able to spatially identify large-scale persistent seasonal changes.

When applied to real MODIS NDVI time series of individual case studies, BFAST Lite and BFASTm-L2 performed particularly well, producing similar performances in detecting the different types of change (**Figure 2.6**). However, on the study area, BFAST Lite in contrast to BFASTm-L2 failed to spatially *highlight* the changes in LULC caused by the setting up of the different types of agro-industrial concessions (LSLAs) (**Figure 2.7**). This is mostly explained by the lower breakpoint magnitudes of BFAST Lite associated with seasonal changes (LSLAs mostly seemed to induce changes in NOS, as shown in **Figure 2.6**), but also because the real background values of BFAST Lite were higher than, the simulated data led us to expect (see Table 2.3). On the contrary, BFASTm-L2 and in a lesser extent Edyn, efficiently spatially capture LSLA driven changes. To better identify the type of change induced by LSLAs, the violin plots of **Figure 2.5** along with the change maps of **Figure 2.7** were helpful in this task. Considering the performance similarities in breakpoint detection between BFAST Lite and BFASTm-L2, and the responsiveness of BFAST Lite magnitudes to abrupt and amplitude changes, in case LSLAs induced changes were of any of these types (i.e. amplitude or abrupt), they would certainly be highlighted in the BFAST Lite's change map. This is not the case. Agro-industrial LSLAs induced changes are therefore probably of seasonal type (NOS and/or LOS). This interpretation is strengthened by the different specific study cases presented in **Figure 2.6**. In addition, because Edyn (which was also able to capture LSLAs changes) outputs breakpoint magnitudes for gradual changes of the same order as the no-change category, the probability that the LSLAs changes captured by BFASTm-L2 but also by Edyn are due to trend changes is low.

In terms of process speed, BFASTm-L2 produced the second-best performance after BFASTmonitor, despite its high rate of false positives, probably enhanced in this study by the absence of a “penalty” period applied after each detected breakpoint, and the non-automatic determination of a stable training period. Indeed, it is worth to remember that the BFASTm-L2 approach relies on the L2-distance, computed whenever a breakpoint is detected by BFASTmonitor. As a result, the speed of this approach depends on the number of breakpoints

found in long time series. With a step frequency of three months, and a monitoring period of 14 years, BFASTm-L2 proved to be faster than the force-brute method represented by L2 computed continuously over areas of more than 100x100 pixels (**Figure 2.8**). Worth is to notice that running times may be improved by changing the step frequency, by removing some of the detected breakpoints if too close one to the other, or by using a more optimized algorithm (that effectively does implement the automatic detection of the training period).

To resume, the main contributions brought to the change detection community through this study are:

- The proposition of BFASTm-L2 as a change detection method faster than BFAST Lite, already known for its speed, making easier its application at larger scales;
- A change detection approach with breakpoint magnitudes more sensitive to seasonal changes. This allows to highlight agricultural-induced LULC changes, supporting the hypothesis that generic LULC changes are very often seasonal, and that more importance should be given to the detection of this type of change.
- More understandable BFASTmonitor, BFAST Lite and Edyn magnitude-based change detection maps. Because of the sensitivity analysis made to the different types of change, insights on the dynamics behind the changes observed in a change map are gained.

### **2.6.2 Recommendations for using BFASTm-L2 to detect and map LULC changes at regional/ national scale**

While it is clear that changes in land use/land cover (LULC) involve changes in the composition of the vegetation which, in turn, result in changes in phenology and seasonality (e.g. a transition from grassland to croplands (Mardian et al., 2021)), the relationship between a change in land use and a change in seasonality is less straightforward. To check this link, we focussed on changes in the system of land use caused by agroindustrial LSLAs installation in the study area. Foreseen specific changes were transitions from natural vegetation or small-scale agriculture to large-scale agricultural systems. These changes were identified without the use of any threshold, mask, or any prior information on the type of land cover or on the direction of the change, meaning the method is suitable for use in unsupervised change detection pipelines. Most of the known LSLAs-related land use changes in the study area were spatially highlighted with BFASTm-L2 and in a lesser extent with Edyn, but not with BFAST Lite, suggesting that these specific LULC changes may occur without necessarily causing any abrupt change.

However, even if BFASTm-L2 produced promising results, some limitations should be mentioned. First, and as pointed out in the violin plots of **Figure 2.5**, the breakpoint highest-magnitude based selection approach hinders the identification of seasonal changes when time series contain abrupt changes. Second, BFASTm-L2 aims to detect seasonal changes based on the use of the time signal canopy greenness (estimated through the NDVI) as a disturbance indicator. In greener environments such as in Casamance (Southern Senegal), where the difference in seasonality may be less marked, the method may be less sensitive to changes in land use. In such cases, other indicators more sensitive to vegetation biomass should be tested. Third, BFASTm-L2 is highly sensitive (i.e. high breakpoint magnitudes) to trends, which are often linked to vegetation recovery/degradation land processes. While trends induced by these land processes are frequently accompanied by changes in amplitude, other seasonal changes such as NOS changes are less expected. Because of this, and because amplitude changes do not profoundly affect the time series shape, shape metrics such as the Procrustes distance could be used in the future to minimize the detection of these type of changes. A last limitation is the inability of BFASTm-L2 to detect recent changes (earlier than the defined L2-w period, in our case 3 years).

When thinking of using this method with other sensors, it is useful to remember that BFASTm-L2 relies on time series that: 1- have a high temporal frequency sufficient to properly represent the phenology, 2- are long enough (8 years as minimum) and 3-are gap-free and smoothed, in order to minimise false detections. As such, applications with Sentinel are currently hampered because of the short temporal depth. On another hand, under tropical conditions and with frequent cloud coverage, highly temporal frequency time series are hardly obtained with Sentinel and Landsat. Finally, because of those sensors' higher spatial resolutions, running times would increase, thus hindering the application of BFASTm-L2 at larger scales. With MODIS data, BFASTm-L2 is able to detect other LULC changes occurring at large scale than those related to LSLAs, such as the changes related to mining and urbanization processes.

Lastly, it is worth emphasising that the method proposed here is purely pixel-based, and that it needs to be completed by a spatial analysis, in order to better identify the drivers of changes in land use and to better interpret the changes. Because of the positive correlation between the NDVI and the breakpoint magnitudes, spatial analysis should be limited to areas that do not encompass many different ecoregions. Further research on spatial analysis at larger scale will be done.

## 2.7 CONCLUSIONS AND PERSPECTIVES

We developed a simple, automatic and rapid approach to select the breakpoint linked to the largest seasonal change in long and dense NDVI MODIS real time series with multiple breakpoints. The method, named BFASTm-L2, is based on the combined use of BFASTmonitor algorithm and the L2 euclidean distance for breakpoint selection, and was shown to accurately detect most of the single change types included in a subsample of the Awty-Carroll et al. (2019) benchmark set. Applied to a study area in Senegal using 20 years of MODIS satellite imagery the algorithm, through the spatialization of its single breakpoint magnitude, proved to be able to spatially identify LULC changes induced by the implementation of agro-industrial concessions in Senegal. This task, performed automatically without the need for any prior knowledge, is fit to be included in unsupervised pipelines to map and analyse generic LULC changes at regional scale. This was also possible because of the absence of any abrupt changes, supporting the hypothesis that generic LULC changes are very often seasonal, and that more importance should be given to the detection of this type of change.

To improve the detection of changes in the LULC at regional and national scales, an operational tool will be developed on a platform such as Google Earth Engine. This will enable supporting land monitoring initiatives such as the Land Matrix in detecting and monitoring anthropogenic changes such as those driven by LSLAs, for which much information remains to be gathered to help ground local teams.

**Acknowledgments:** The authors wish to thank the ISRA-BAME and the Land Matrix Initiative for data gathering and support. This work was supported by the French Space Agency (CNES) through the funding of the TOSCA-VISAGE project. Y. Ngadi received a fellowship from Montpellier University. This work was also supported by the French National Research Agency under the Investments for the Future Program #DigitAg referred as ANR-16-CONV-0004.

**Author Contributions:** Yasmine Ngadi processed the data, performed the proposed methods, analyzed the results and wrote the paper; Mohamadou Dieye and Jérémy Bourgoïn retrieve and processed the field data base; Agnès Bégué, Valentine Lebourgeois and Anne-Elisabeth Laques supervised the research and contributed to the editing and review of the paper.

**Conflicts of Interest:** The authors declare no conflict of interest.

# 3 INSIGHT INTO LARGE-SCALE LULC CHANGES AND THEIR DRIVERS THROUGH BREAKPOINT CHARACTERIZATION – AN APPLICATION TO SENEGAL

---

This chapter is based on:

Ngadi Scarpetta, Y., Lebourgeois, V., Laques, A.-E., Dieye, M., Bégué, A., 2024. Insight into large-scale LULC changes and their drivers through breakpoint characterization – An application to Senegal. *Int. J. Appl. Earth Obs. Geoinf.*. *Under review*



### 3.1 HIGHLIGHTS

- BFASTm-L2 change characterization provides insight into national land dynamics
- Differentiation between change types improves understanding of LULC change drivers
- The RGB change map is useful for inferring main drivers of LULC change (LULCC)
- Time series shape dissimilarity is sensitive to agricultural-driven LULC changes
- Large-Scale Agricultural Investments driven LULCC are unsupervisedly detectable

### 3.1 ABSTRACT

As global land cover/ land use change (LULCC) threatens the human's well-being, accurate detection and characterization of LULCC is of paramount importance. The increasing availability of dense satellite image time series (SITS), together with the ever-improving change detection algorithms, *has allowed significant progress to be made. However,* much remains to be done in its characterization.

This study aims to provide insight into the land dynamics and drivers of change in Senegal, through an RGB composite change map based on BFASTm-L2 detected changes and three NDVI time series-derived change metrics. These metrics, chosen to discriminate different types of change, include: the magnitude of change, the direction of change, and the time series shape dissimilarity. The sensitivity of each metric to different types of change was first tested on a simulated dataset, and then applied to MODIS NDVI SITS (2000-2021). The RGB change map allowed the visualization of different "signatures" of change, which, in combination with ground information, rainfall data, NDVI time series analysis and Google Earth imagery, helped to link them to different drivers of change. Climatic and anthropogenic changes, such as those induced by LSAI or mining, could be visually inferred from the RGB map.

Although only tested in Senegal, this study shows the usefulness of integrating the type of change, especially seasonal change, into the characterization of land change. This approach, has the advantage of being fast, interpretable, robust to noise and easily transferable to different regions.

## 3.2 INTRODUCTION

The Earth's land surface has been changing at an unprecedented rate, with about three-quarters of the land surface having been modified by humans within the last millennium (Winkler et al., 2021). Because such global land changes threaten the sustainability of ecosystem services and human's well-being, there is a strong requirement for monitoring land cover and land use changes (LULCC) (Radwan et al., 2021). Over the past few decades, the land change community has benefited from the rapid advances in remote sensing technologies, together with the free and open data policy, cloud computing platforms, and the ever-improving change detection algorithms. The great potential of dense Satellite Image Time Series (SITS) analysis have triggered a paradigm shift from bi-temporal change detection to continuous monitoring of LULC change, fostering interest in land use mapping and monitoring, especially at regional and higher scales (Molinier et al., 2021; Weiss et al., 2020; Woodcock et al., 2020; Zhu et al., 2022).

Among change detection algorithms, the trend is toward those that can use all available data (and thus handle seasonal variations) in dense SITS, with supervised approaches (map classification, trajectory classification), unsupervised statistical approaches, ensemble approaches and recently, deep-learning approaches (Molinier et al., 2021). Unsupervised statistical methods, if fast enough and with few tuning parameters, are suitable for large-scale studies where the availability of labeled data at appropriate spatial and temporal resolution remains a major challenge (Woodcock et al., 2020). While they have been widely adopted by the land change community through cloud-based platforms, they still have limitations.

First, fast algorithms that do not rely on time series decomposition are more sensitive to abrupt and long-term gradual changes than to seasonal changes due to the use of harmonic regression models. This can be problematic if the goal is to detect seasonal changes. In fact, specific land-use conversions, such as those driven by Large Scale Agricultural Investments (LSAIs), often include seasonal changes without abrupt changes (Ngadi Scarpetta et al., 2023). Second, algorithms that perform time series decomposition are often too computationally expensive to be applied on a large scale. To the best of our knowledge, none of these statistical techniques are currently capable of both detecting and identifying the specific *type of seasonal change*, i.e., in amplitude, in the number of seasons (NOS) and/or in the length of season (LOS).

In an attempt to fill this gap, Ngadi Scarpetta et al. (2023) proposed BFASTm-L2, a rapid change detection approach, fine-tuned to seasonal changes. The algorithm demonstrated higher sensitivity to NOS changes than three state-of-the-art algorithms, allowing spatial detection of LULCC induced by LSAIs in Senegal. This and other studies (e.g. Hentze et al.

(2017), Mardian et al. (2021)) highlight the importance of accurate seasonal change detection, but also the particular link that may exist between the type and the driver of change, allowing to shift from land change detection to characterization, which remains one of the most difficult challenges in the land change community (Verburg et al., 2009; Zhu et al., 2022). In a recent review of the potential of remote sensing to fully characterize land change, Zhu et al. (2022) proposed a multifaceted framework consisting of five facets, including *Where* (i.e. the location of change), *When* (i.e. the date of change), *What* (i.e., target of change), *How* (i.e., the metrics of change), and *Why* (i.e., the drivers of change). They found that while the first three facets have been studied extensively, much work remains to be done on the last two facets.

Given the identified needs, our main objective is to contribute to the *How* and *Why* facets of land change using the BFASTm-L2 algorithm. While its sensitivity to seasonal changes was demonstrated, BFASTm-L2 also showed some sensitivity to abrupt changes and trends, hindering the analysis of land dynamics from a change type perspective. Therefore, this study aims to: i) derive a set of change metrics with varying sensitivities to the different types of change, and ii) combine these metrics into a comprehensive RGB change map to provide insight into the drivers of change at the national scale. Compared to other change visualization approaches (Hird et al., 2016; Julien and Sobrino, 2021), this approach goes beyond detecting change to characterizing it (changes are detected using BFASTm-L2), by looking at the specific relationships between the different types of change (particularly seasonal) and potential drivers. Characterized changes are the most important ones occurring at each location within the entire monitoring period. Particular attention was given to LSAs. Details of the approach, data and methods are provided in the next section.

### 3.3 DATA AND METHODS

#### 3.3.1 Approach

Flowchart of the approach is shown in **Figure 3.1**. The first step aims to characterize BFASTm-L2 detected change on MODIS dense NDVI SITS. In addition to the magnitude of change, two change metrics were derived: a time series shape dissimilarity measure to assess the type of seasonal change, and an NDVI change ratio to assess the change’s directionality. The second step aims to provide a map of the major drivers of change on a national scale by combining the change metrics into a unique RGB change map. Dominant colors were tentatively assigned to one or a few drivers of change by visual inspection using Google Earth (GE), NDVI time series analysis, precipitation distribution analysis, and the LSAI field database.

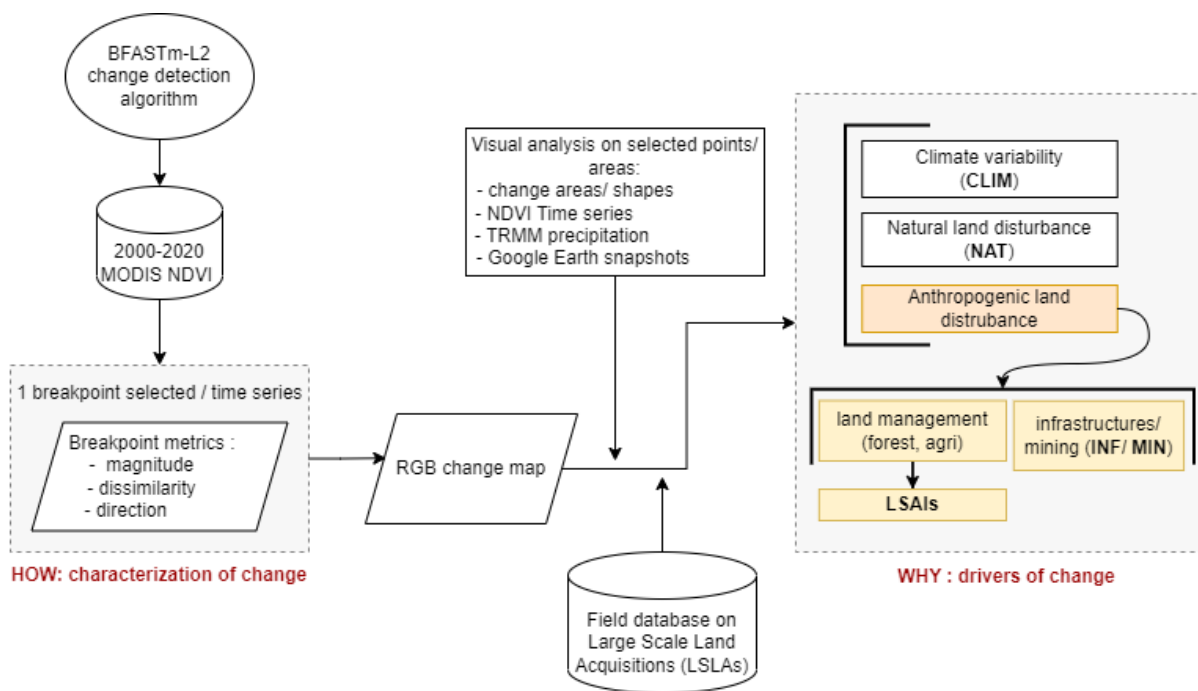


Figure 3.1: Flowchart for identifying drivers of change from MODIS time series.

#### 3.3.2 Study area and LSAI

Located in the westernmost part of the Sahel, Senegal has a strong north-south rainfall gradient resulting in a semi-arid climate in the north (200-400 mm/ year) and a tropical climate in the south (800-1200 mm/ year) (**Figure 3.2**). Senegal has two distinct climatic seasons: a dry season from November to May and a rainy season from June to October, with the main land cover types being steppe, savanna and sub-humid dry forest (Budde et al., 2004; Sultan and Janicot, 2003; Tappan et al., 2004). It also has remarkable ecosystems, such as the productive wetlands along the Senegal River, which have supported small farmers, herders,

fishermen and traders for centuries, but are increasingly threatened by dams and irrigated rice schemes (Horowitz and Salem-Murdock, 1993; Tappan et al., 2004).

The agriculture, which accounts for 15% of the GDP, is dominated by smallholdings with farm sizes of less than 5 hectares (Bourgoin et al., 2019). Large Scale Land Acquisitions (LSLAs) are however increasing in number, with 3% of the country's arable land declared under contract by foreign investors in 2016 (Harding et al., 2016). Due to opacity, lack of geospatial information, and potential socio-environmental impacts, efforts have been made to inventory and map LSLAs (Bourgoin et al., 2019; Nolte et al., 2016). However, discrepancies and gaps remain due to the spatio-temporal dynamic nature of LSLAs and differences in methodologies. Automated and rapid approaches are needed to easily monitor the entire national territory.

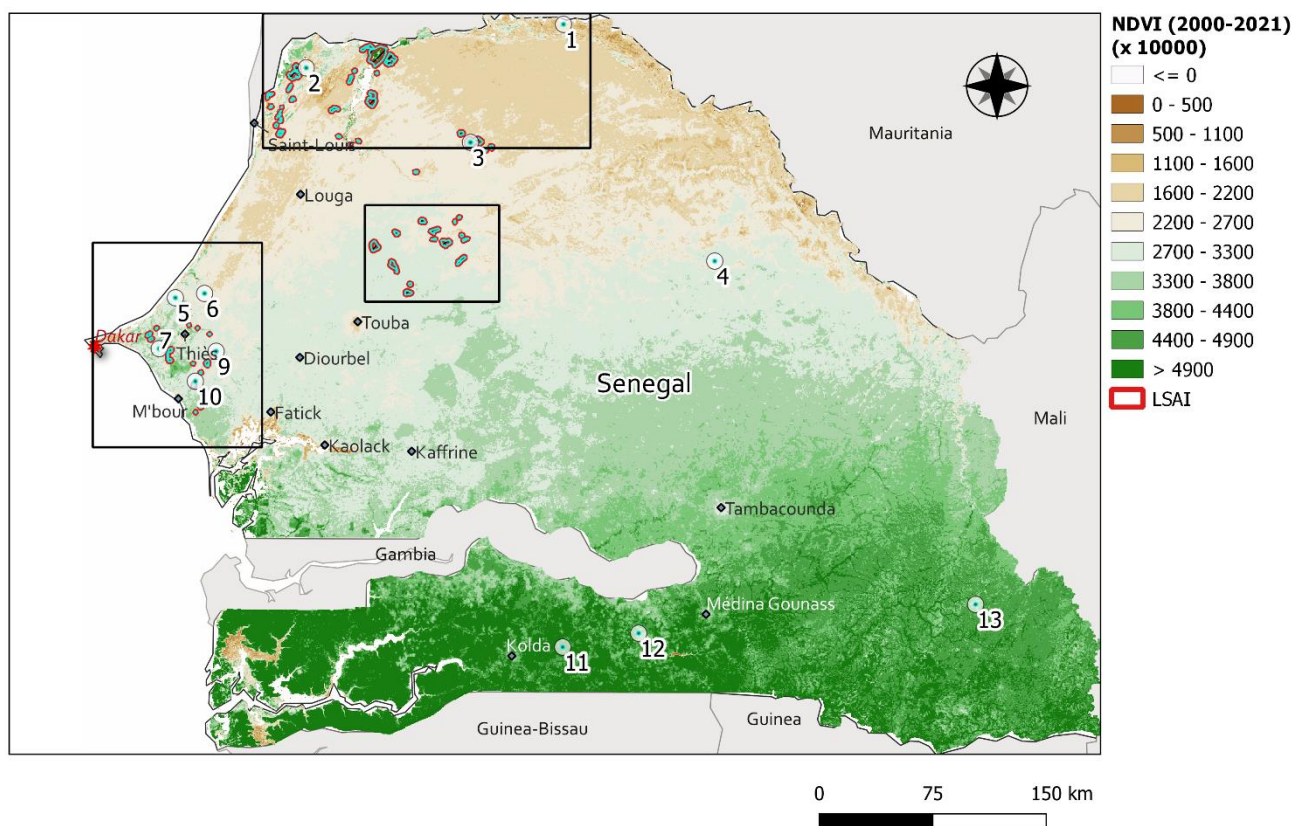


Figure 3.2: Senegal's map of the MODIS NDVI 2000-2021 average. The black boxes represent the three regions of interest: a) the Senegal river (SR) (North), b) the Niayes (West), and c) Ferlo (Centre) that include most of the LSLAs (red polygons) reported in the field database (M. Dieye, personal communication, 2022). Punctual study cases used in this study are represented by points 1 to 13.

Agricultural LSLAs, hereinafter refer as to **Large Scale Agricultural Investments or LSAs**, are mainly concentrated over three regions, shown in **Figure 3.2** (Bourgoin et al., 2019):

- 1- The Senegal River (hereinafter refer as to SR) region shown in box a, is an important agricultural region with a growing number of LSAs, mainly focused on horticulture, sugarcane production and cereals, mainly rice.

- 2- The box b in the Niayes includes many LSAs dedicated to horticulture. The vegetation consists mostly of open agricultural parkland.
- 3- The sylvopastoral area of Ferlo shown in box c, consists mainly of tree and shrub savannah, and is home to most of the LSAs focusing on gum arabic production.

**Figure 3.2** also displays the locations of thirteen study cases with medium-large magnitudes of change, corresponding to different LULCCs, as interpreted using GE imagery, LSAI field database, and analysis of selected NDVI time series and rainfall distributions.

### 3.3.1 Data

#### 3.3.1.1 LSAI dataset

In 2019, the Senegalese Institute of Agricultural Research (ISRA) conducted a field campaign on LSLAs, with more than 700 polygons recorded in a spatial database (M. Dieye, personal communication, 2022). The database contains deal information, such as deal type (agrobusiness, mining, etc), size, year of transaction/ negotiation or implementation status. A sub-database of 76 polygons was used, which only includes: i) LSAs established or expanded during the monitoring period (2003-2018), ii) with a minimum size of 30 hectares, iii) with at least 1/3 productive area (as verified by GE imagery).

#### 3.3.1.2 Simulated time series dataset

In this study a gap-free and noise-free subsample of the simulated time series dataset created by Awty-Carroll (2019) (<https://osf.io/taf9y/>) was used to analyze the sensitivity of the proposed metrics to different types of change. The subsample consisted of 3,150 simulated NDVI time series (2006-2015 at 16-day frequency resolution as MODIS), each containing a single change and belonging to one of the following change types: trend, abrupt with/without a trend change, amplitude, length of season (LOS), and number of seasons (NOS). Different intensities of change, with 50 replicates for each, were included in each group. The highest absolute intensities are (Fig. 3): 0.3 NDVI units for the amplitude changes (+ 60% of the initial amplitude), a shift of season start of -45 days backward for LOS changes, + 1 season for NOS changes, 0.046 NDVI units/ year for trend changes, and + 0.3 of the NDVI baseline (+ a trend of 0.046 NDVI units/year) for abrupt changes. More information (on the parameters, intensities...) can be found in Awty-Carroll et al. (2019).

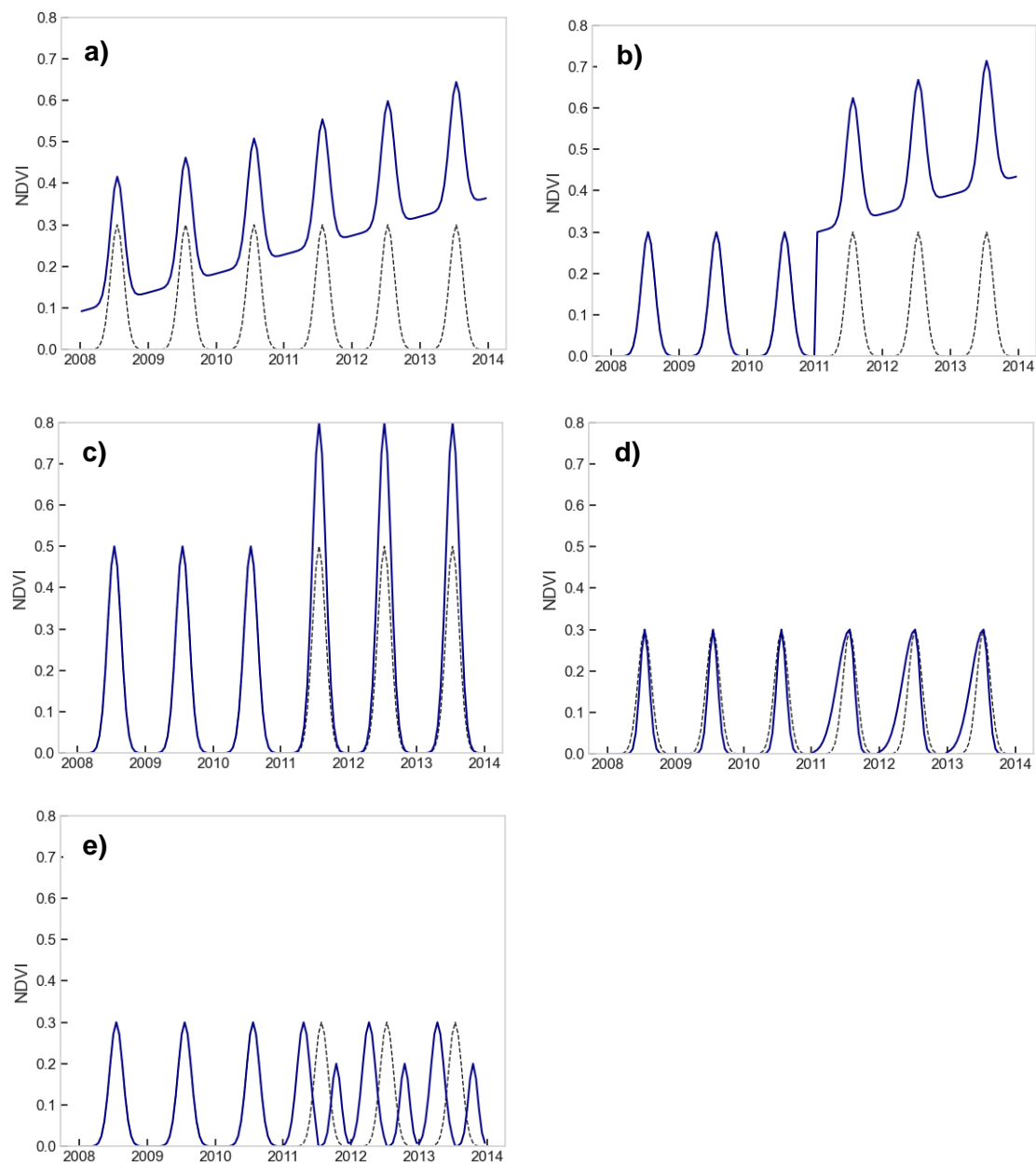


Figure 3.3: Illustration of the highest-intensities changes found in the simulated dataset, for each change-type group: a) trend (gradual change), b) abrupt change with a gradual change, c) amplitude change, d) change in the length of season (LOS), e) change in the number of season (NOS). The no-change time series is plotted in black dash line.

### 3.3.1.3 MODIS NDVI data and pre-processing

The Moderate Resolution Imaging Spectroradiometer (MODIS) is a satellite sensor launched in 1999, and designed to improve our understanding of global dynamics and processes on Earth. Its global coverage, moderate spatial resolution (250 m) and high temporal resolution (1-2 days), make it ideal for detecting subtle land cover changes. Here, a set of MODIS NDVI 16-day composites at 250 m resolution (MOD13Q1, Collection 6), was acquired for Senegal over 2000-2021 and pre-processed in GE Engine. Pre-processing included the application of an optimized weighted Savitzky-Golay smoothing (Chen et al., 2004). Weights were computed according to Piou et al. (2013b), which uses the reliability of the pixel (i.e., quality flag, view



zenith angle) and the position of each observation in a predefined moving window (exponentially decreasing weights). A moving window length of 13 observations and a polynomial order of 3 were used.

### 3.3.1.4 TRMM precipitation data

Precipitation estimates for the study cases 1 (16.605°; -14.627°), 4 (15.220°; -13.708°) and 11 (12.944°; -14.631°) were obtained from the Tropical Rainfall Measuring Mission (TRMM), a satellite designed to observe rainfall in tropical and subtropical regions of the world (Kummerow et al., 1998). Specifically, the 3B43v7 product, created using TRMM-adjusted data from several sources (namely high-quality microwave data, infrared data, and rain gauges analysis), was downloaded from GE Engine. Monthly precipitation rate estimates (mm/hr monthly average) at a spatial resolution of 0.25° were converted to annual estimates (mm/year).

## 3.3.1 Methods

This section is divided into two sub-sections. Section 2.4.1, dedicated to the *How* facet of land change, presents the three-change metrics derived from the MODIS NDVI SITS that characterize in different ways the change found by BFASTm-L2 between 2003 and 2018. Section 2.4.2, dedicated to the *Why* facet of land change, presents the RGB map obtained from the combination of the three-change metrics, enabling the identification of possible drivers of change through the analysis of known study cases. Throughout this study, special emphasis was placed on the detection of LSAs.

### 3.3.1.1 Contributing to the How facet of land change: the change metrics

#### 3.3.1.1.1 The magnitude of change metric

The magnitude of change used in this study is the one corresponding to the largest magnitude breakpoint detected by BFASTm-L2 in each time series (Ngadi Scarpetta et al., 2023). This magnitude represents the Euclidean distance (i.e., the square root of the sum of the squared difference) between the two 3-year time series located at each part of the breakpoint. Because this distance does not take into account the non-stationarity of the variance in the time series (Lhermitte et al., 2011), it is very sensitive to trends and amplitude changes that are commonly attributed to natural (e.g., forest regeneration) and climatic variability-induced changes, respectively. To discriminate such contributions, a time series shape similarity metric expected to be more sensitive to seasonal changes is introduced in the next section.

#### 3.3.1.1.2 The time series shape dissimilarity metric

Since special attention is given to the detection of LSAs in Senegal, that induce changes in NOS (i.e. number of seasons) and LOS (length of season) (Ngadi Scarpetta et al., 2023), we propose here the Procrustes distance (hereafter referred to as dissimilarity) as a method to refine the type of seasonal change detected. Procrustes analysis is a statistical shape analysis



that optimally superimposes (by translation, rotation, reflection and scaling) two (or more) vectors (Gower, 1975). Values  $\in [0,1]$ , with higher values indicating greater dissimilarity. In this study, the Procrustes function of the Python spatial package *scipy* was used.

As for the magnitude in our previous study, a sensitivity analysis of the dissimilarity metric was first performed on the simulated time series dataset and on a selection of six real NDVI time series for pixels with no change and pixels undergoing different types of change. For each sample, the dissimilarity between the two 3-year time series (first averaged on a monthly basis) at each part of the change was computed.

### 3.3.1.1.3 The change direction metric

To assess the direction of change, the ratio of the 3-year NDVI average after the change to the 3-year NDVI average before the change was calculated. An NDVI ratio below 1 indicates a “negative” direction of change, while a ratio above 1 indicates a “positive” direction of change.

As for the magnitude in our previous study, a sensitivity analysis of the NDVI ratio to different types of change was performed on the simulated time series dataset and on a selection of real NDVI time series.

### 3.3.1.2 *Contributing to the Why facet of land change: combining the change metrics into an RGB composite map*

Although there is no one-to-one relationship between the drivers and the types of change, some drivers are more likely to cause a particular type of change. First, climate variability often causes direct changes in amplitude through its effects on vegetation vigor and health. If important enough to change the existing vegetation cover, it may also affect LOS (Evans and Geerken, 2006). Second, abrupt changes (often accompanied by amplitude changes) are often associated with large abiotic (fires, floods...) or anthropogenic changes. When it comes to gradual changes, land management practices inducing subtle changes such as selective logging, reforestation, or biotic changes (forest regeneration, disease...) may be in cause. Finally, drivers likely to induce changes in LOS and NOS include agricultural activities (particularly agricultural intensification), that have a direct impact on the land cover type (Arvor et al., 2012; Brown et al., 2007; Hentze et al., 2017; Ngadi Scarpetta et al., 2023).

In light of these observations, we propose a composite RGB map based on three change metrics with different sensitivities to different types of change: the magnitude of change in red, the NDVI ratio (indicative of the direction of change) in green, and the dissimilarity measure (sensitive to seasonal changes) in blue. The dominant change “signatures” (i.e. colors) observed in the RGB map were linked to specific land dynamics and drivers of change (see Table 3.1) based on the: i) expected relationships between some drivers and types of change (presented above), ii) visual verification of thirteen selected study case points presenting

different land use transitions, iii) constructed knowledge of LSAs, and iv) comparison of NDVI time series and the distribution of annual average precipitation for cases that are likely to be non-anthropogenic due to their area and shape (cases 1, 4, and 11). As BFASTm-L2 breakpoint detection is based on the comparison of two 3-year time series subsamples, only potential drivers with short-term effects were considered and grouped into four broad classes: CLIM (for changes induced by climate variability), NAT (for biotic natural changes), MIN/INF (for mining/ infrastructure), and LSAI (for intensive agricultural activities), the latter being our main focus in this study.

### 3.4 RESULTS AND DISCUSSION

#### 3.4.1 Sensitivity assessment of the dissimilarity and NDVI ratio metrics

The sensitivity of the dissimilarity and NDVI ratio metrics is here assessed using two different datasets.

##### 3.4.1.1 On the simulated dataset

###### 3.4.1.1.1 The dissimilarity metric

Table 3.1 shows the dissimilarity medians computed for each type of change in the simulated dataset. Null values are observed for "vertical" changes, represented by changes in amplitude, trend, and abrupt changes (with/without trend changes). In contrast, it is very sensitive to NOS changes (median = 0.78) and to a lesser extent to LOS changes (median = 0.35).

Table 3.1: Medians of the dissimilarities computed per type of change using the simulated dataset.

	Amplitude	LOS	NOS	Trend only	Abrupt w/ trend	No change
<b>Dissimilarity</b>	0.03	0.35	0.78	0.03	0.04	0.03

###### 3.4.1.1.2 The NDVI ratio metric

**Figure 3.4** shows the distribution of the NDVI ratios for each type of change in the simulated dataset. This metric shows particular sensitivity to abrupt changes (max. = 6.03), and to a lesser extent to LOS changes (max. = 1.85). Less sensitivity is shown for amplitude and gradual changes (max. of 1.62 and 1.59 respectively). The lowest value is reached for NOS changes (1.47).

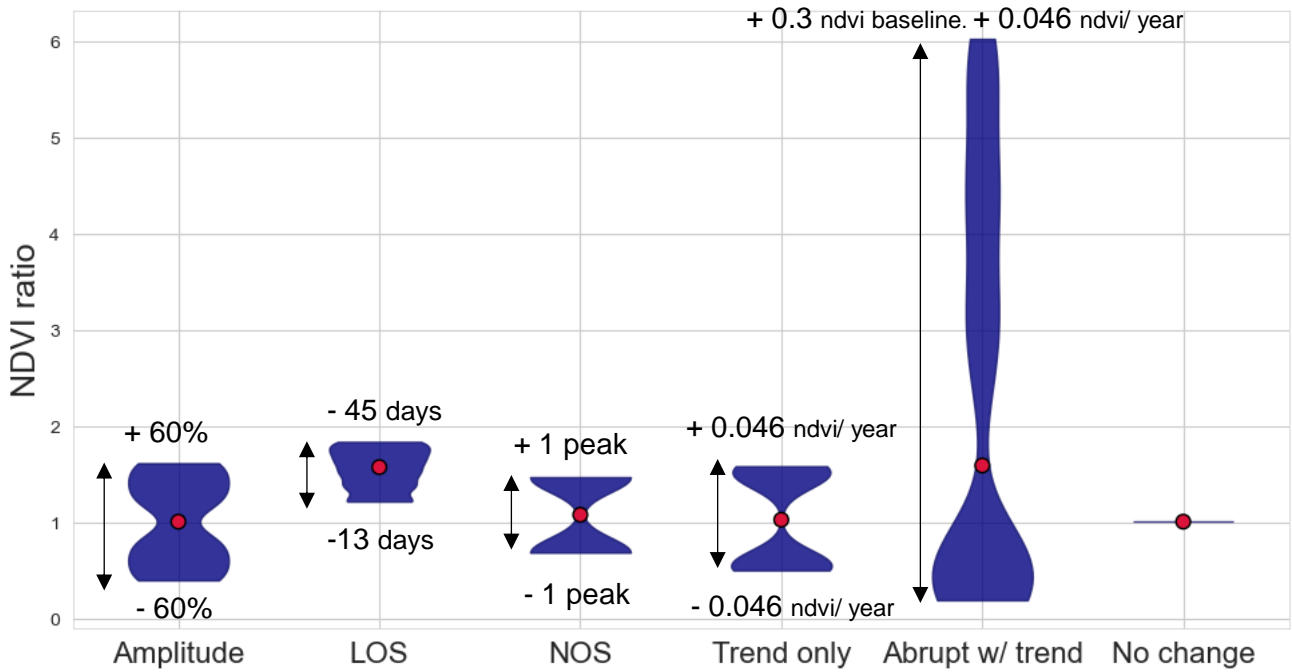


Figure 3.4: Violin plots of the NDVI ratio for each type of change in the simulated dataset. Group's medians are represented by red dots.

### 3.4.1.2 On real NDVI time series with different land use transitions

In this section, the MODIS NDVI time series of six pixel-study cases are presented. Four of them include LULC transitions from/to (Fig. 5.a-d): natural vegetation other than estuaries/wetlands (NAT), estuaries (EST), small-scale agriculture (SA), large-scale agricultural investment (LSAI), mining (MIN) and infrastructures (INF) (roads or airports). The computed dissimilarities and NDVI ratios are given.

## Chapter 3: Results and Discussion

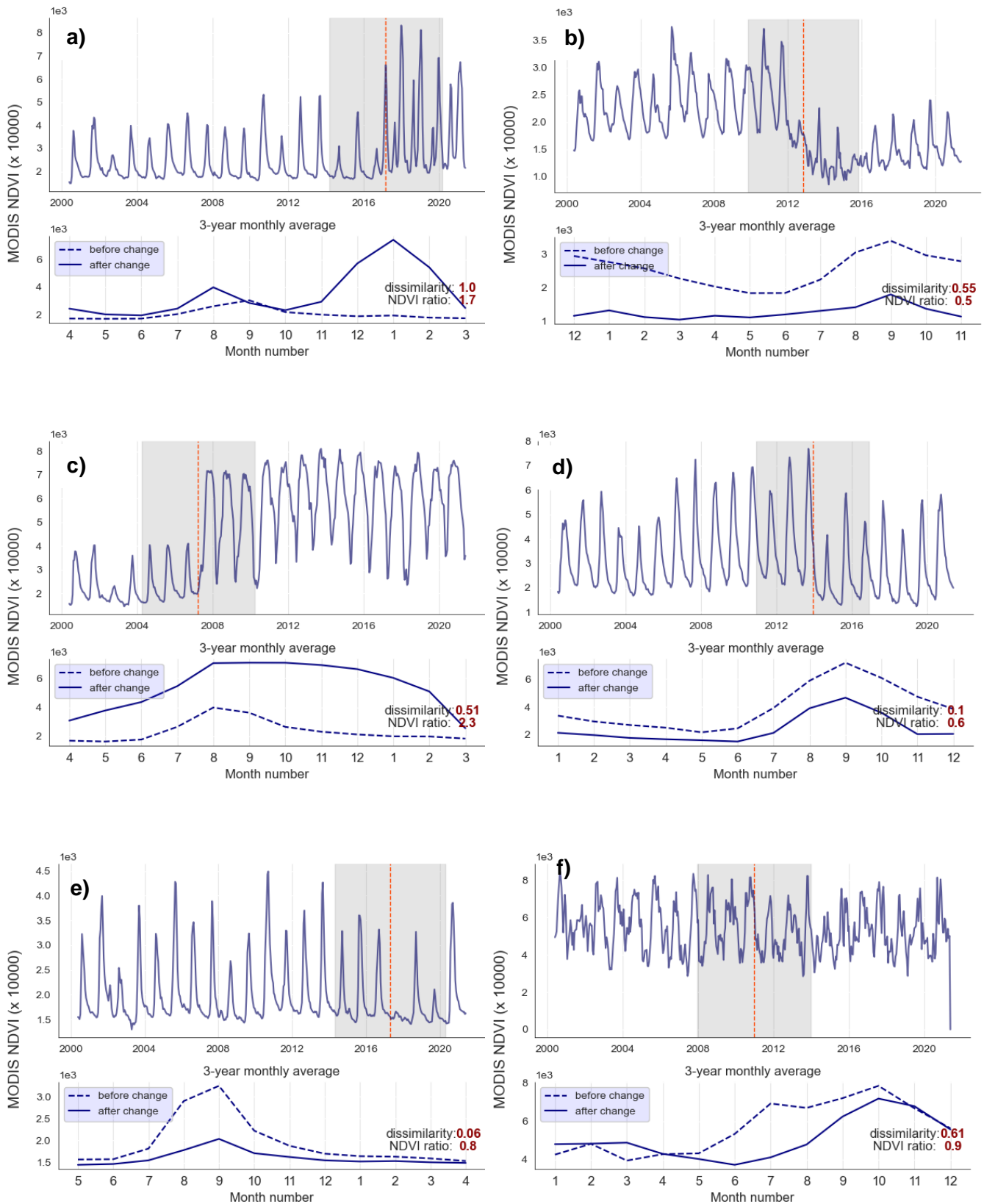


Figure 3.5: Selected pixel-study cases. Each subplot's upper part shows the smoothed MODIS NDVI time series, with the breakpoint detected by BFASTm-L2 (red dashed line), and the 3-year time period before and after the breakpoint (grey zone). The bottom part shows the 3-year monthly average (before/after the breakpoint) and the computed change metric values. Land transitions and coordinates [latitude, longitude] are: a) NAT-LSAI [16.181°

-15.779°]; b) SA-MINE [15.039°, -16.806°]; c) NAT-LSAI [16.406°, -15.689°]; d) NAT-INF [14.708°, -17.090°]; e) NAT [16.278°, -15.312°]; f) EST [12.838°, -16.384°].

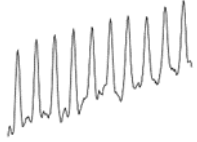
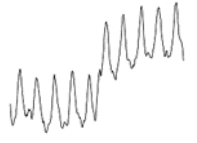
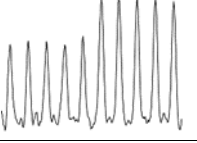
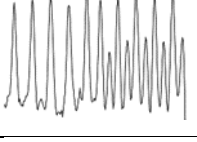
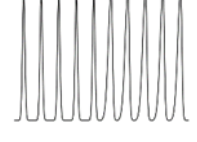
**Figure 3.5** demonstrates that real data can be complex, with various types of changes occurring simultaneously. For example, two study cases with the same land use transition (NAT-LSAI) show different combinations of change types. While case a. shows a combination of amplitude (~ +75%) and NOS changes, case c. shows a combination of amplitude, LOS and abrupt change. Transition SA-MINE study case b) shows a combination of all types of changes. Study cases without land conversions, i.e. cases e. and f. (NAT and EST), show mainly amplitude changes.

In this dataset the highest and lowest NDVI ratios were obtained for NAT-LSAI (case c: 2.3) and SA-MINE (case b: 0.5) transitions, due to large abrupt (~ -0.1 NDVI units) and amplitude (+75%) changes, respectively. At the opposite, the NDVI ratios close to 1 of the two cases without land conversions (case e: 0.8; case f: 0.9) indicate the absence of significant land changes.

In terms of dissimilarity, values greater than 0.5 were observed for large-scale land use conversions, such as conversion to LSAI (case a: 1; case c: 0.51) or mining (case b: 0.55). Cases without LULCC (e.g. climate-induced changes: case e), or those covering a smaller area (e.g. infrastructure construction: case d), have lower dissimilarity values (0.06 and 0.1). An exception is the estuary study where a high dissimilarity is observed (case f: 0.61) due to the water level fluctuations.

The results are summarized in Table 3.2. Results related to the magnitude sensitivity were taken from a previous study (Ngadi Scarpetta et al., 2023).

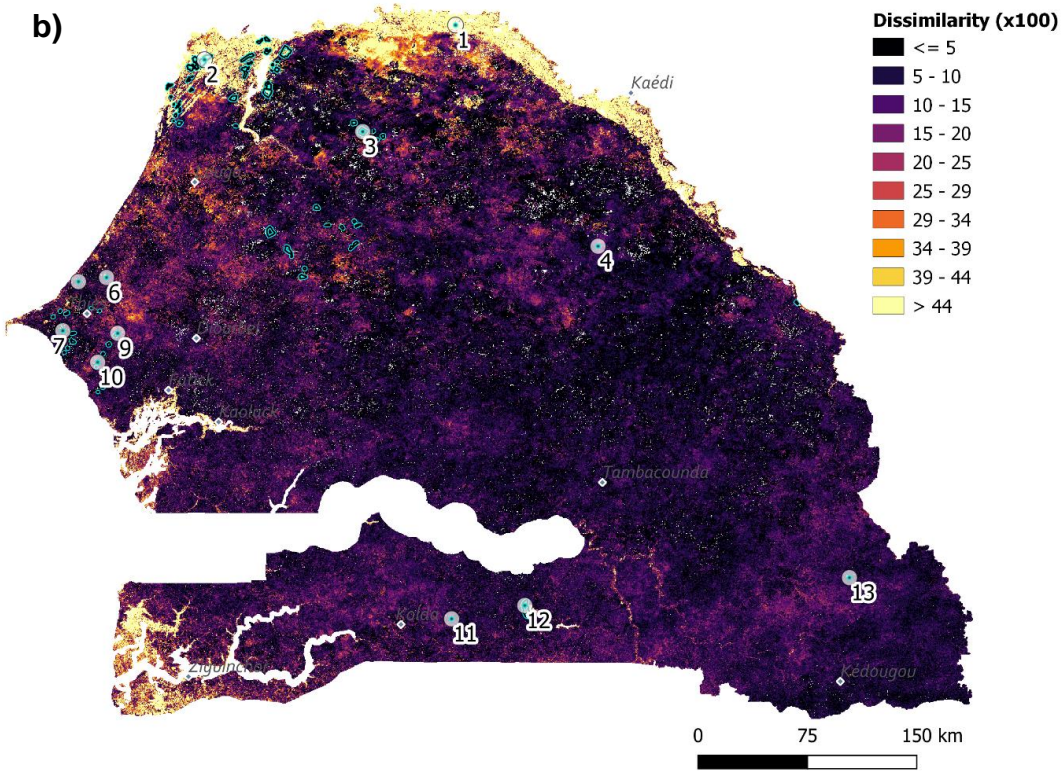
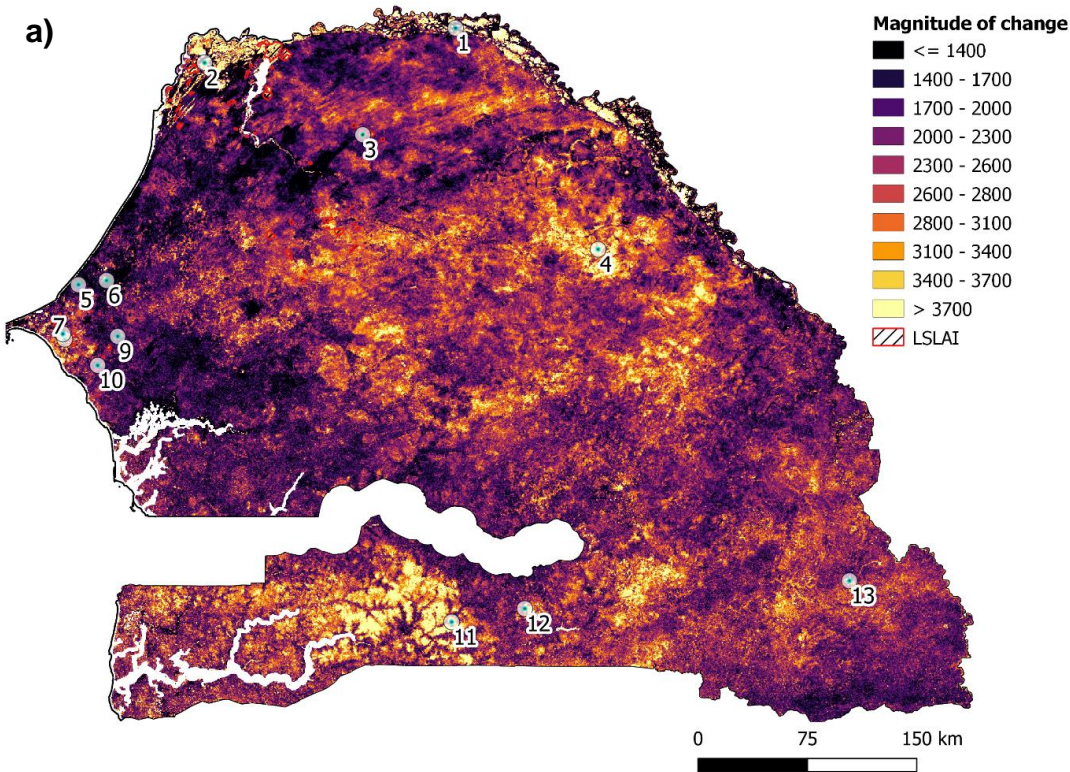
Table 3.2: Sensitivity of the change metrics to different types of change. Sensitivity classes are: low: +, medium: ++, high: +++. None: - Change type thumbnails are for illustration only, as changes can be positive/ negative, from one direction to the other.

	Change metric	Magnitude	Dissimilarity	NDVI ratio
<b>Change type</b>				
<b>Trend</b>		++	-	++
<b>Abrupt</b>		+++	-	+++
<b>Amplitude</b>		++	-	++
<b>NOS</b>		+++	+++	+
<b>LOS</b>		+	++	++

3.4.2 Contributing to the How facet of land change

Figure 3.6 shows the different change metrics maps at the national scale: magnitude of change (Fig. 3.6.a), time series shape dissimilarity (Fig. 3.6.b), and NDVI ratio (Fig. 3.6.c). To facilitate readability, close-in views of these maps for the three regions (SR, Niayes and Ferlo) are given in Appendices B to D. As a reminder, the changes are those detected by BFASTm-L2 in the full MODIS NDVI time series between 2003 and 2018. Several observations can be made from these maps.





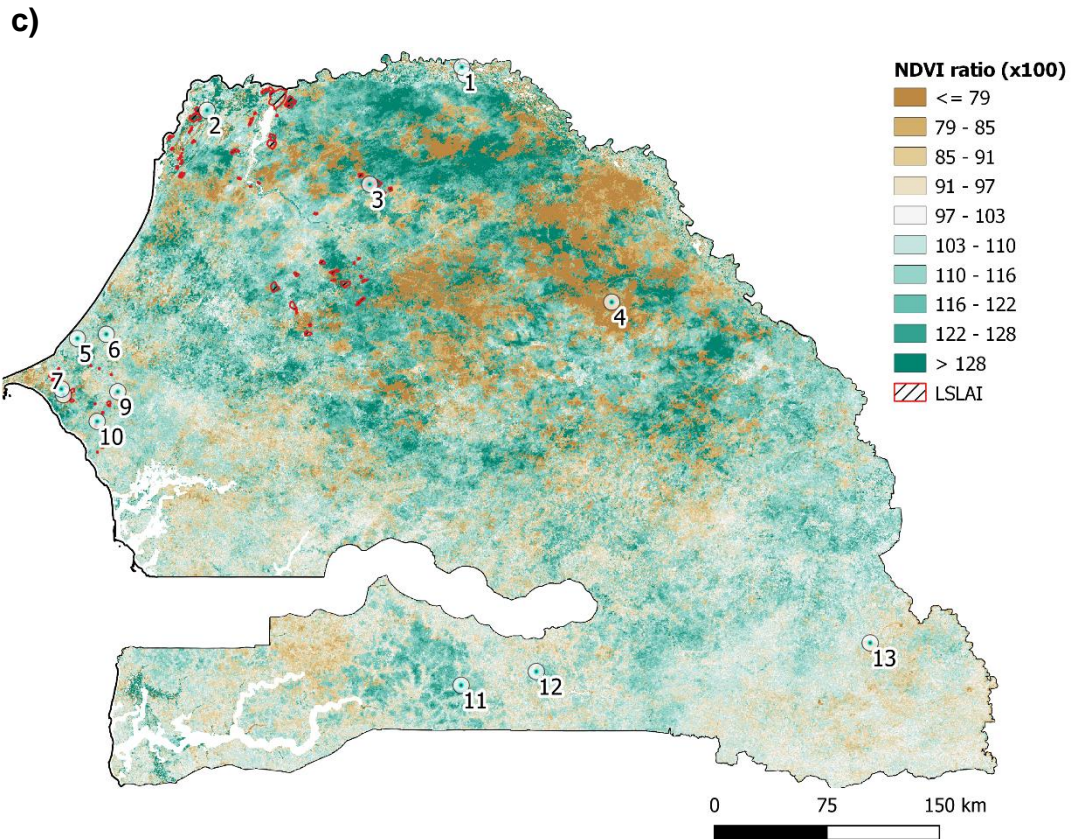


Figure 3.6: National maps of the different change metrics: a) the magnitude, b) the dissimilarity, c) the NDVI ratio.

The first is that a breakpoint is almost always found on long and dense SITS (see Appendix A for the date of change map).

Second, the magnitude of change map (Fig. 6.a) highlights significant areas of change throughout the monitoring period. Large-scale changes (big yellow patches) are primarily concentrated in the eastern north and central pastoral regions, and the forested areas in Casamance, which are located around point 4 and point 11 respectively. The changes occurred almost simultaneously in each region (Appendix A), indicating a common cause of change per region. At a smaller scale, it appears that the highest magnitudes in the SR (Appendix 3.B.c) are linked with agricultural activities or wetlands, which sharply contrast with the remaining arid environment. In the Niayes (Appendix C.c), the LSAIs are well highlighted with large magnitudes of change. Additional highlighted structures include linear structures (points 7-8) that correspond to infrastructure constructions, compact patches (point 6) that correspond to mines, and more diffuse patterns over greener areas (Appendix C.a). In the Ferlo (Appendix D.c), random patterns of high magnitudes and different dates of change appear. Here, the LSAIs dedicated exclusively to the production of gum arabic do not induce visible changes.



Third, most of the previously observed magnitude hotspots disappear in the dissimilarity map (**Figure 3.6.b**). In this map, high values are particularly observed in the SR valley, near water bodies and in the estuary regions in the south, indicating high seasonal variations. Elsewhere, the values are low, but with minor local variations. In the Niayes (Appendix C.d), the LSAs are particularly highlighted by the dissimilarity metric. In the SR region, the LSAs also show high dissimilarities, although they are not clearly distinguished from the surroundings characterized by the presence of wetlands. Conversely, the Ferlo region (Appendix D.d) exhibits very low dissimilarity values.

Finally, the NDVI ratio map enables evaluation of change direction (“negative” for values between 0-1, and “positive” for values >1) and demonstrates overall similar patterns to the magnitude of change map. Very high (around 1.3) or very low (below 0.8) NDVI ratios are associated with high magnitudes of change. The most significant changes in NDVI occur in the eastern pastoral region, with a combination of high and low values (Appendix D.e). Some patterns are also observed in Casamance (southwest region), but with fewer extremes, except for the forested region near point 11 that displays high NDVI ratios. At a smaller scale, the LSAs in the SR region are generally well represented by compacted positive change patches (point 2 and polygons in Appendix B.e), with the exception of LSAs established prior to 2003 (Appendix 3.B.f). In this region, the wetlands have moderate positive or significant negative NDVI ratios, suggesting that these ecosystems have mostly dried out (all causes confounded) during the monitoring period. In the Niayes (Appendix C.f), the linear structures observed in the magnitude of change map and attributed to infrastructure are well highlighted by negative values (points 7-8 of **Figure 3.8**), as well as the mines (Fig. 8.6). In this region, the LSAs do not have overall high positive NDVI ratios, indicating less abrupt changes than in the north. The highest positive values observed here are in the form of diffuse patterns, associated with greening areas.

As observed, each map provides useful information on its own, but it is necessary to consider all three maps together for a full understanding of the changes. To this end, in the next section we propose an RGB composite map constructed to highlight anthropogenic changes.

**3.4.3 Gaining insights on the *Why* facet of land change: the RGB composite map** **Figure 3.7** shows the national-scale RGB composite map, constructed from the BFASTm-L2 magnitude of change (red band), the NDVI ratio (green band), and the dissimilarity metric (blue band), as well as the 2000-2019 MODIS NDVI and TRMM annual precipitation distribution for 3 pixels located in large natural areas (forest: point 11, wetland: point 1, other natural vegetation: point 4). This figure is followed by close-in views over the 13 study cases (**Figure 3.8**), to help interpret the observed signatures in terms of drivers of change.

The dominant colour observed is green, followed by yellow and orange, indicating changes with low dissimilarities. Green indicates small magnitude changes, along with varying NDVI ratio values that range from 0.73-1.33. Orange and yellow pixels indicate high magnitudes of change with either a decrease in NDVI average (orange pixels) or an increase in NDVI average (yellow pixels). Based on their size and irregular shape, as well as their similar dates of change (see Appendix A), it is likely that these large orange and yellow patches are caused by natural or climatic drivers of change. To confirm this hypothesis, we visually analyzed the NDVI time series and precipitation distributions from points 4 and 11 in **Figure 3.7**. The change detected at point 4 (red dashed line in 2012) correlates with a decrease in both NDVI amplitude and precipitation, supporting the hypothesis of a climate-driven change. In fact, a drought episode that caused a major humanitarian crisis in the Sahel was reported in 2012 (United Nations Office for the Coordination of Humanitarian Affairs). The NDVI time series at point 11 shows a sharp increase in its baseline in 2003 (detected date of change), corresponding to an increase in pluviometry. This is consistent with positive anomalies observed in the region during the same period (Solly et al., 2020). The absence of significant seasonality changes, other than erratic amplitude changes, supports the assumption of non-anthropogenic change.

Looking closely at **Figure 3.8** and the sub-regional areas in Appendices B, C, additional shades of pink and blue appear that are associated with changes of greater dissimilarity values (along with varying magnitudes). The blue shades are primarily observed in the northern and southern coastal ecosystems (see Appendix 3.B.f). These areas show small magnitudes of change and no significant NDVI changes over a 3-year period. In the arid north, areas of light pink are associated with higher magnitudes of change (Fig. 8.1), suggesting a higher degree of instability compared to estuaries in the south.

In addition to coastal ecosystems, other land dynamics are shown in light pink. This is the case for all LSAs in the Niayes and several of those in the SR (zooms 2, 5, 7-8, 9, 10 of **Figure 3.8** and Appendices B.f and C.f). However, those dedicated to gum arabic production in the Ferlo do not follow the same trend (in yellow, see Appendix D and **Figure 3.8.3**). In the SR, many of the LSAs established before 2003 (start of the monitoring period) are colored in light pink, while the most recent ones appear in white, indicating very high NDVI ratios (zoom 2 of Figure 3.8). In the southern Casamance, the LSAI plots shown **Figure 3.8.12** are light pink, except for the most recent plot in orange, indicating very low NDVI ratios, possibly due to vegetation removal associated with plot preparation.

As noted above, the orange color indicates changes with low dissimilarity values. It is therefore surprising that the vegetation cover's removal does not lead to higher dissimilarities. This is also the case for most changes induced by infrastructure (roads Fig. 8.8) and some mines (Fig.

8.9), which are also orange. Larger mines (Fig. 8.6, 8.13) or infrastructure (airport in Fig. 8.7) tend however to appear in dark pink, indicating higher dissimilarities.

Chapter 2 · Results and Discussion

Dissimilarity

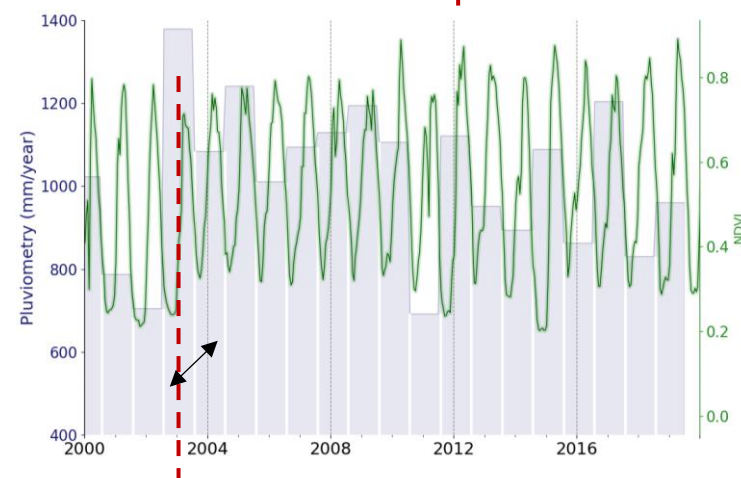
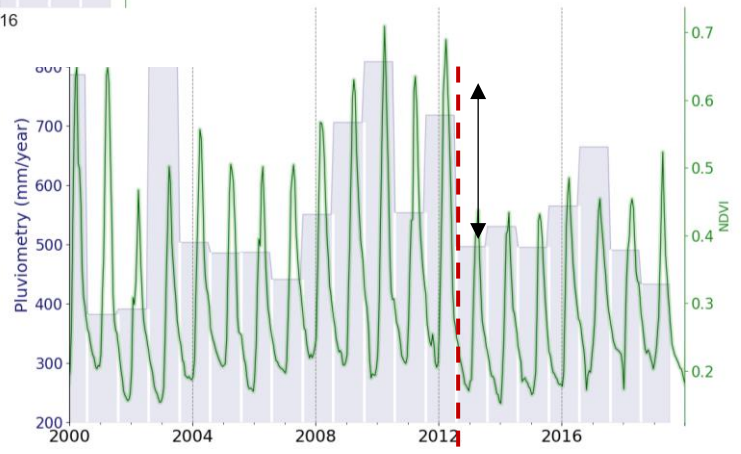
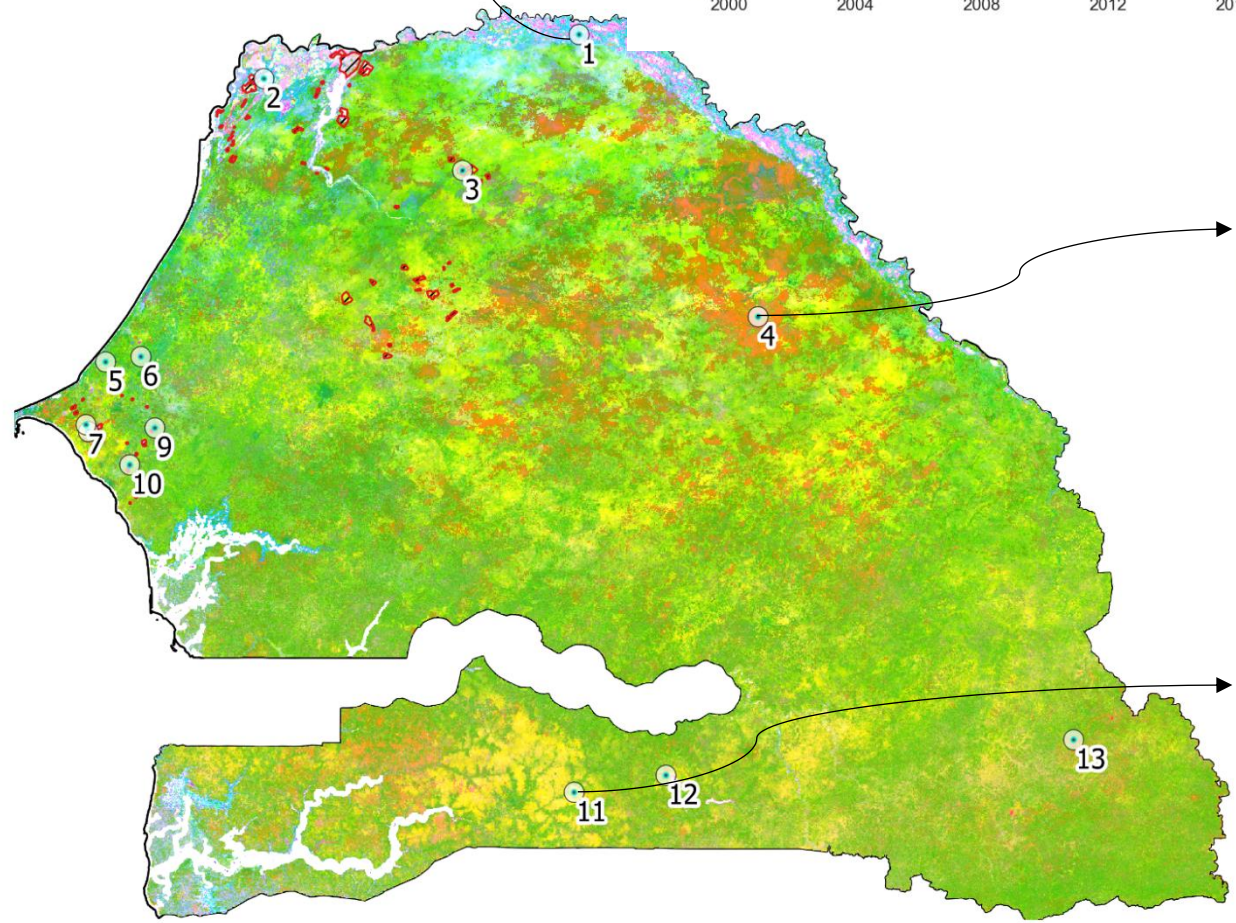
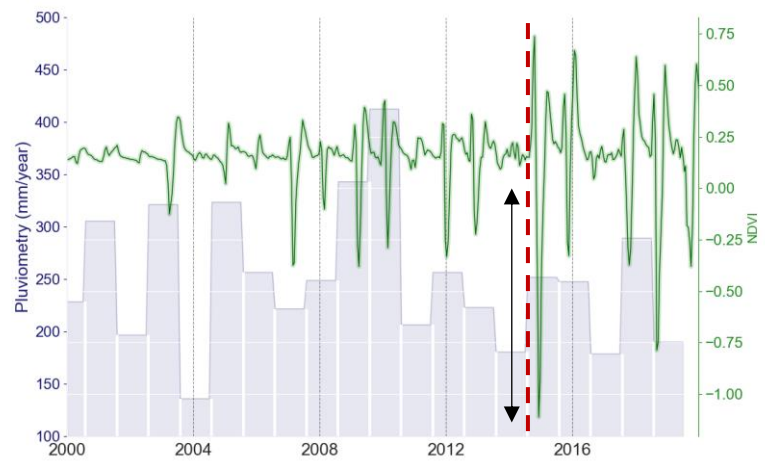
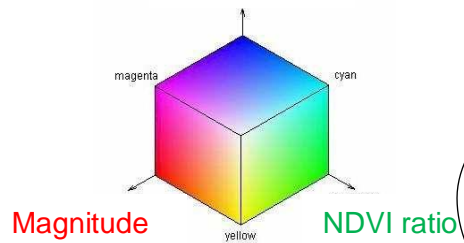


Figure 3.7: RGB composite map with in Red: the change's magnitude, in Green: the NDVI ratio, and in Blue: the dissimilarity metric. MODIS 2000-2019 NDVI and TRMM annual rainfall distribution are shown for three pixels with natural/ climate-driven changes : 1 (wetland), 4 (shrub savanna) and 11 (dry tropical forest) . See legend in Table 3. Map values were stretched between the 1st and 99th data percentiles, corresponding to [0.73-1.33] and [0-0.79] for NDVI ratio and dissimilarity, respectively.



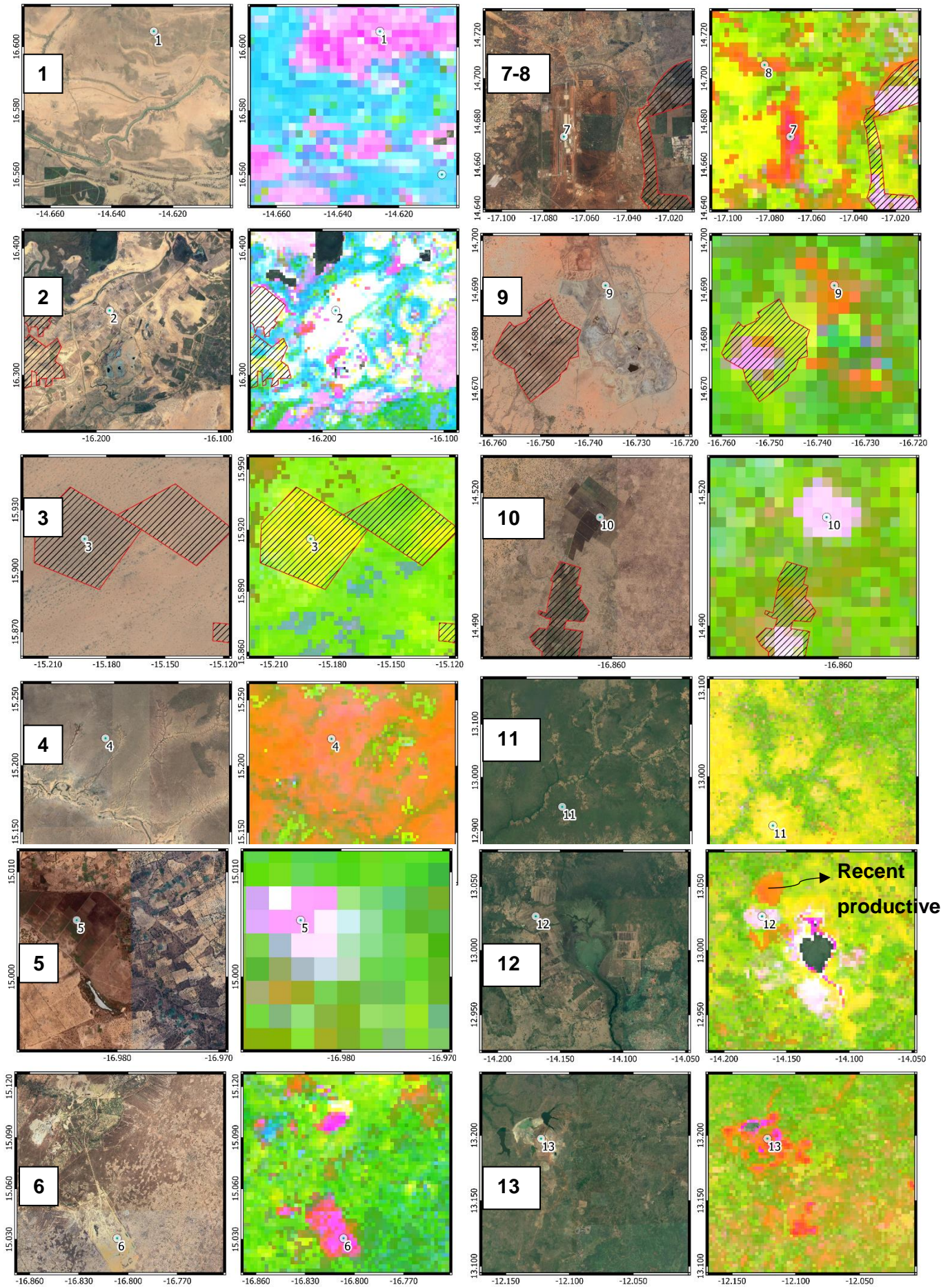


Figure 3.8: Zoom-ins of study cases 1 to 13. LSAIs: 2, 3, 9, 10, 12. Not in database LSAIs: points 5, 10. Mines: points 6 and 13. Infrastructures: points 7 (airport) and 8 (road). Forest: point 11. Wetland: point 1. Other natural vegetation: point 4.

Chapter 3 : Results and Discussion

The observations are summarized in Table 3.3. In this table, the dominant colors observed are linked to the 3-band intensities (1st column), and to some potential drivers of change (2nd column).

Table 3.3: Association table between the RGB map colours (first column) and the potential drivers of change (2<sup>nd</sup> column, see 2.4.2). Change signatures are composed of, in red: the magnitude of change, in green: the NDVI ratio and in blue: the dissimilarity metric.

Band			Possible driver of change
R:	G:	B:	
Magnitude	NDVI ratio	Dissimilarity	
-	↓ = +	+	CLIM
-	↓ = +	-	NAT, CLIM
+	+	-	NAT, LSAI
+	+	+	LSAI
+	=	+	LSAI, CLIM
+	-	-	MIN/INFR, CLIM, (LSAIs in tropical ecoregions)
+	-	+	MIN/INFR

## 3.5 DISCUSSION

### 3.5.1 The *How* facet

Detected changes are usually categorized as either abrupt or gradual changes depending on the duration of the change (Zhu et al., 2022). Sudden seasonal changes have not been explicitly addressed because: the detection of changes in dense SITS often requires the removal of the seasonality (Evans and Geerken, 2004; Hird et al., 2016), they are mostly considered to be climate-driven changes (especially for erratic changes in amplitude), and they typically result in small magnitudes of change (especially for LOS/NOS changes) (Ngadi Scarpetta et al., 2023). Because land use conversions, such as LSAI-driven ones in Senegal, typically involve seasonal changes that are not necessarily accompanied by abrupt changes, the effort in this study was to select metrics that effectively discriminate between types of changes

The first change metric is the BFASTm-L2 magnitude of change, which is known to be sensitive to abrupt and large gradual changes, but also to NOS changes (Ngadi Scarpetta et al., 2023). Because many change drivers are likely to induce these types of changes, this map represented our baseline map, from which different drivers, all of which inducing high-intensity changes, were tentatively discriminated.

The second metric evaluated here was the dissimilarity metric introduced to help discriminate seasonal changes. This metric showed to be invariant to amplitude and trend changes, while being very sensitive to NOS/LOS changes and small intra-annual variability (Table 3.1 and **Figure 3.5**). This metric was most effective in the Niayes, where it effectively highlighted LSAIs (Appendix C). In other regions, the dissimilarity shows a great sensitivity to the strong intra-annual variability present in the wetlands and estuaries, in the SR floodplain (Appendix 3.B.d) and Casamance respectively. Although this metric shows a strong sensitivity to agricultural changes, it seems to be primarily sensitive to the seasonal changes caused by annual crops, rather than slow-growing plantations, as is the case with the arabica gum plantations in Ferlo (Appendix D).

Finally, the NDVI ratio based on the comparison of two 3-year periods after and before the change, provide an indication of the direction and intensity of the change. A 3-year period was considered sufficient to favor persistent changes over climate-driven changes. While this ratio improves the discrimination between biomass-producing and biomass-depleting drivers, we found that it was most useful for discriminating anthropogenic drivers such as mining or infrastructure (zooms 7-9,13 in **Figure 3.8**), rather than LSAIs, which have highly variable NDVI ratios (Appendices B.e and C.e).



While each metric alone provides valuable insight into potential drivers of change, interpretation of these multiple changes independently is often complicated.

### 3.5.2 The *Why* facet

#### 3.5.2.1 *General considerations*

The RGB map based on the magnitude of change, the dissimilarity and the direction of change represents a simple change visualization method, providing clues as to the possible drivers of change. While the change detection is done at the pixel level, the attribution of possible drivers of change is done by looking at the color (“signature” of change), area, and shape of the clusters. The date of change map also helps in this process, as very large clusters of change with the same date of change are unlikely to have an anthropogenic origin. This is the case of the natural areas around points 4 and 11 in **Figure 3.7**, which are most likely driven by pluviometry and natural forest regrowth respectively. These represent the most significant events occurring at any given time during the 2003-2018 monitoring period. While current dynamics may be slightly different, this change map is powerful for detecting punctual human-induced events in the past. Compared to natural and climate-driven changes, anthropogenic land changes are spatially constrained, often with geometric shapes.

#### 3.5.2.2 *LSAIs*

While the LSAIs show different signatures of change across the country, most share high dissimilarities with different NDVI ratios, introducing a new way to characterize them. All LSAI-related changes appear in light pink/white, except for those in the Ferlo, which are yellow (very low dissimilarities). Specifically, LSAIs are best identified in the Niayes (Appendix C), where they are widely spaced and are coloured light pink (Figs. 8.5, 8.10, 8.11). In this region no other land dynamics produce the same signature of change. In the arid north (Appendix 3.B), newly established LSAIs appear in white as a result of significant changes that represent land conversion from semi-arid natural vegetation to agribusiness. LSAIs established prior to the monitoring period and SR Valley wetlands characterized by flood recession agriculture both appear in light pink, making them difficult to distinguish. Finally, in the semitropical south, new LSAIs (or spatial extensions, i.e. the orange area in Fig. 8.12) show negative abrupt changes that may be related to field preparation and biomass (woody vegetation) removal in favor of annual crops. In contrast to the north, plots of LSAIs installed before 2003 appear in very pale pink, indicating changes in agricultural practices. The presence of these white clusters near compact orange clusters may help distinguish LSAIs from other land dynamics in the tropics, even though in this region, as in the Niayes, no other land dynamics appear to produce the same colors.

Compared to other pixel-based studies aimed at detecting LSAI (Bey et al., 2020; Hentze et al., 2017; Humi et al., 2017; Xiao et al., 2020), this approach has the advantage of being fast,



unsupervised, not crop-specific, independent of absolute spectral values, and able to use all available data. Therefore, it can be used as the first step of a pipeline to detect potential LSAs. A more in-depth analysis could then be performed using HR satellite imagery only at these specific locations, including morphological and textural metrics following Vogels et al. (2019). Because the detected changes represent the largest LULCC within the entire monitoring period, detected areas may have a different land use in the present.

### 3.5.3 Limitations and Recommendations

Our results show that the RGB change-metric based map at the national scale proved to be useful for quick visual detection of specific land changes. However, there are some limitations. The first, as shown in Table 3.3, is that there is not a one-to-one relationship between a given combination of change types and a change agent. For example, while it is true that the reported Senegalese LSAs are likely to cause seasonal NOS/LOS changes, other spatial objects such as the highly unstable estuaries and wetlands are also exhibit this type of change. In this case, other characteristics such as spatial patterns (area, shape) may help to determine the most likely driver of change. Similarly, the same type of change's agent can cause different combinations of change types. For example, certain types of LSAs, such as those involving slow-growing tree plantations in semi-arid environments, do not cause visible NOS/LOS changes such as annual crops. These land dynamics are therefore poorly captured by BFASTm-L2. Finally, it is important to note that the results may be different in other climatic regions, especially in the humid tropics, where the observed seasonal changes may be less pronounced. The same conclusions apply to regions where the diversity and size of agricultural systems makes LSAI less contrasted with other agricultural land uses. Indeed, it is important to keep in mind that in Senegal, the majority of the farms are smallholdings, rain-fed, with an area of less than five hectares (Bourgoin et al., 2019).

Regarding some of the recommendations, it is worth noting that because the change magnitude is not NDVI normalized, for a same region encompassing different biomes (e.g. the Niayes), the magnitude tends to be higher over the forested areas. On the other hand, when considering the use of this method with other sensors, because BFASTm-L2 relies on long (at least 8 years) gap-free and smoothed time series, with a high temporal frequency to properly represent phenology, applications with higher resolution SITS such as Sentinel are currently hampered by the short temporal depth. The use of coarse resolution MODIS SITS allows rapid and easy identification of areas with specific land change dynamics over large areas, which can be analysed in more detail using HR satellite imagery at a later stage.

### 3.6 CONCLUSIONS AND PERSPECTIVES

In this exploratory study, the BFASTm-L2 change detection algorithm was applied to MODIS 2000-2020 NDVI imagery to provide insights into the major land changes and potential drivers of change in Senegal, contributing in a novel way to the How and Why facets of land change proposed by Zhu et al. (2022). The How facet of land change was characterized here by three change metrics, namely the magnitude of change, the direction of change, and a time series shape dissimilarity metric. The combination of these metrics in an RGB composite map allowed the characterization of different land dynamics, and proved to be a useful visualization approach in detecting different anthropogenic LULCC such as those induced by LSAIs, mines or infrastructure. Complex land use systems such as LSAIs, which are diverse in terms of cropping practices, are often difficult to detect. However, by combining the “signature” of change, with other change characteristics such as the area and shape, newly installed (within the monitoring period) LSAIs could be visually inferred from the RGB map.

Although this approach has only been tested in Senegal, it demonstrates the usefulness of integrating the type of change, and in particular the seasonal ones, into the characterization of land change. This approach, based on a statistical change detection method, has the advantage of being interpretable, robust to noise and easily transferable to different regions, as it uses all the available temporal data and does not require the use of ancillary data. Further research will focus on automating the LSAI detection approach and integrating morphological and textural variables from high spatial resolution satellite imagery into the analysis. The approach will then be tested to different regions of the world.

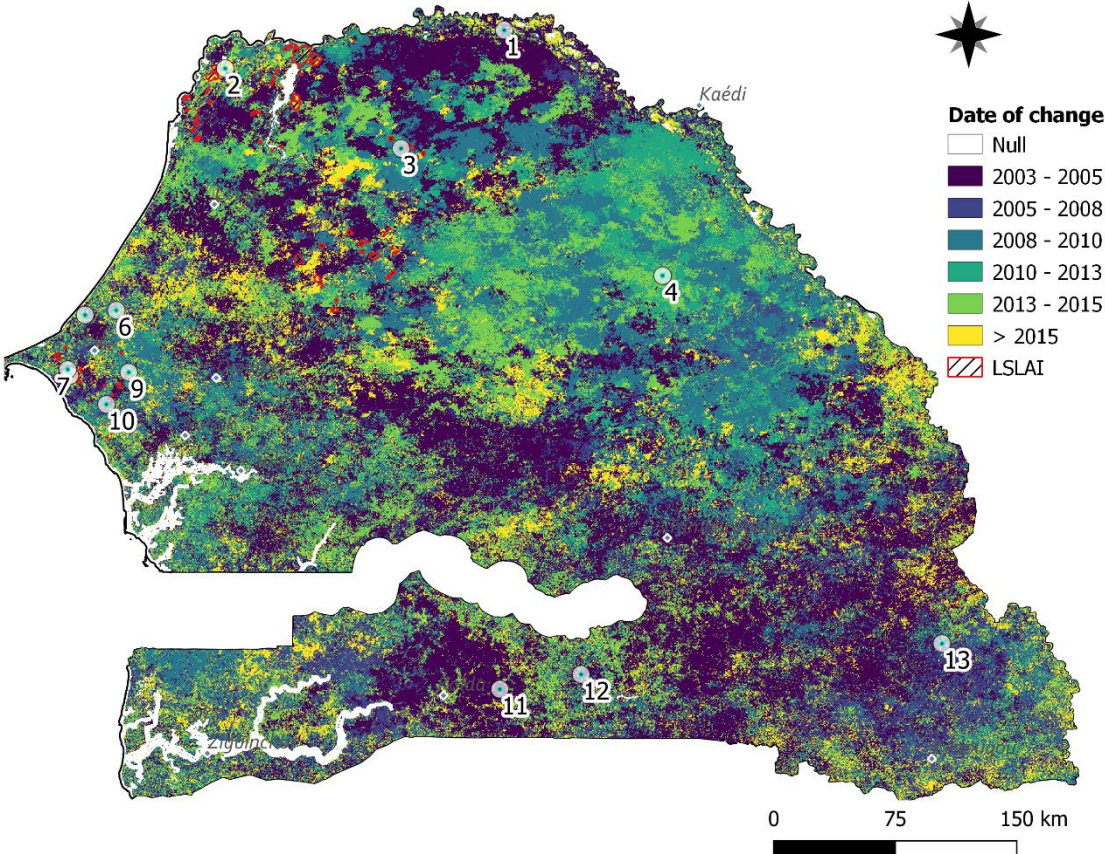
**Acknowledgments:** The authors thank the Land Matrix Initiative and the ISRA-BAME for data collection and support. This work was supported by the French Space Agency (CNES) (TOSCA-VISAGE project), and by the French National Research Agency under the Investments for the Future program #DigitAg (ANR-16-CONV-0004). Yasmine Ngadi received a scholarship from the University of Montpellier.

**Author Contributions:** Yasmine Ngadi processed the data, proposed and implemented the proposed methods, analyzed the results and wrote the paper; Mohamadou Dieye retrieved and processed the field database; Agnès Bégué, Valentine Lebourgeois and Anne-Elisabeth Laques supervised the research and contributed to the editing and review of the paper.

**Conflicts of Interest:** The authors declare no conflict of interest.

### 3.2 APPENDICES

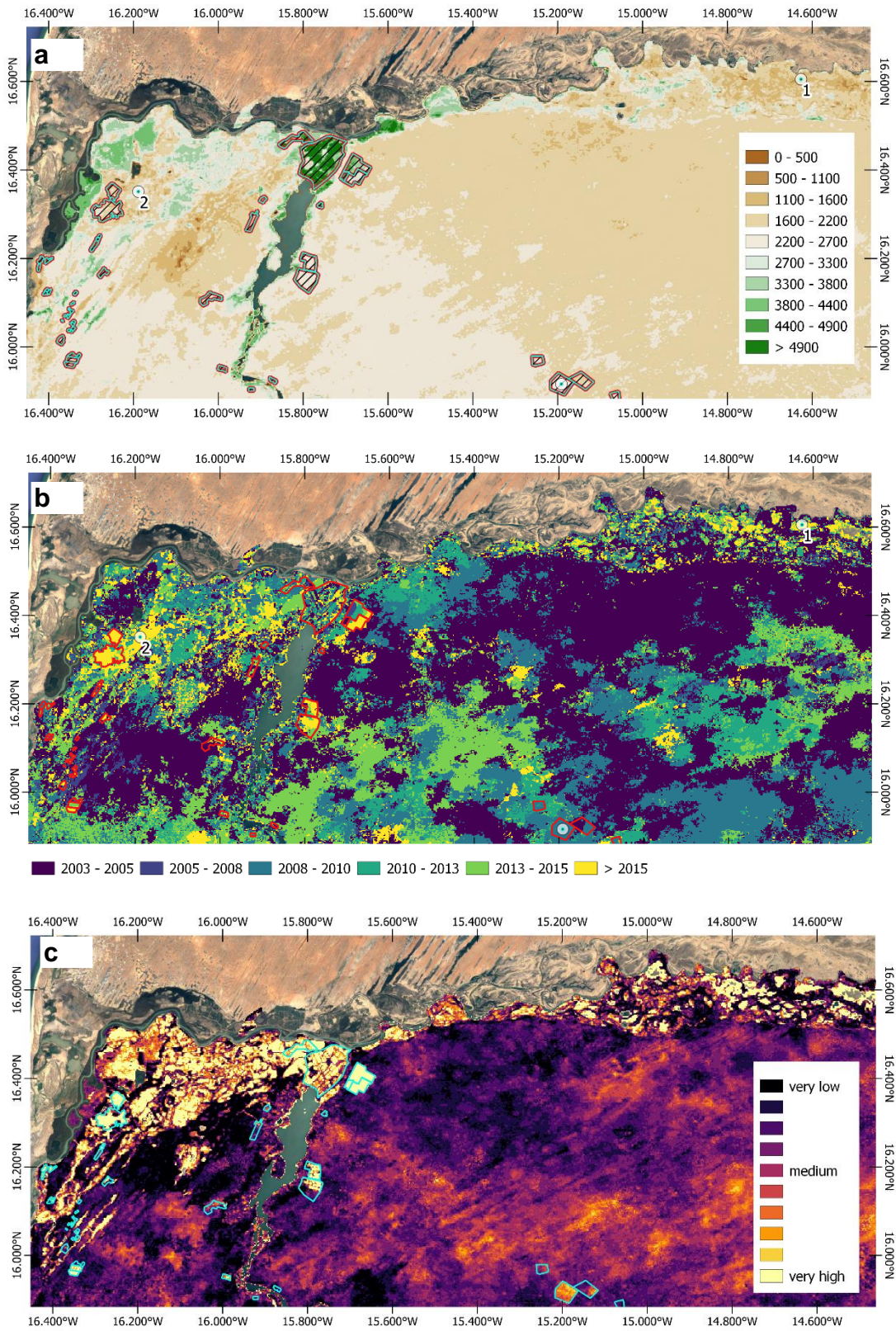
#### Appendix 3.A



Date of change map

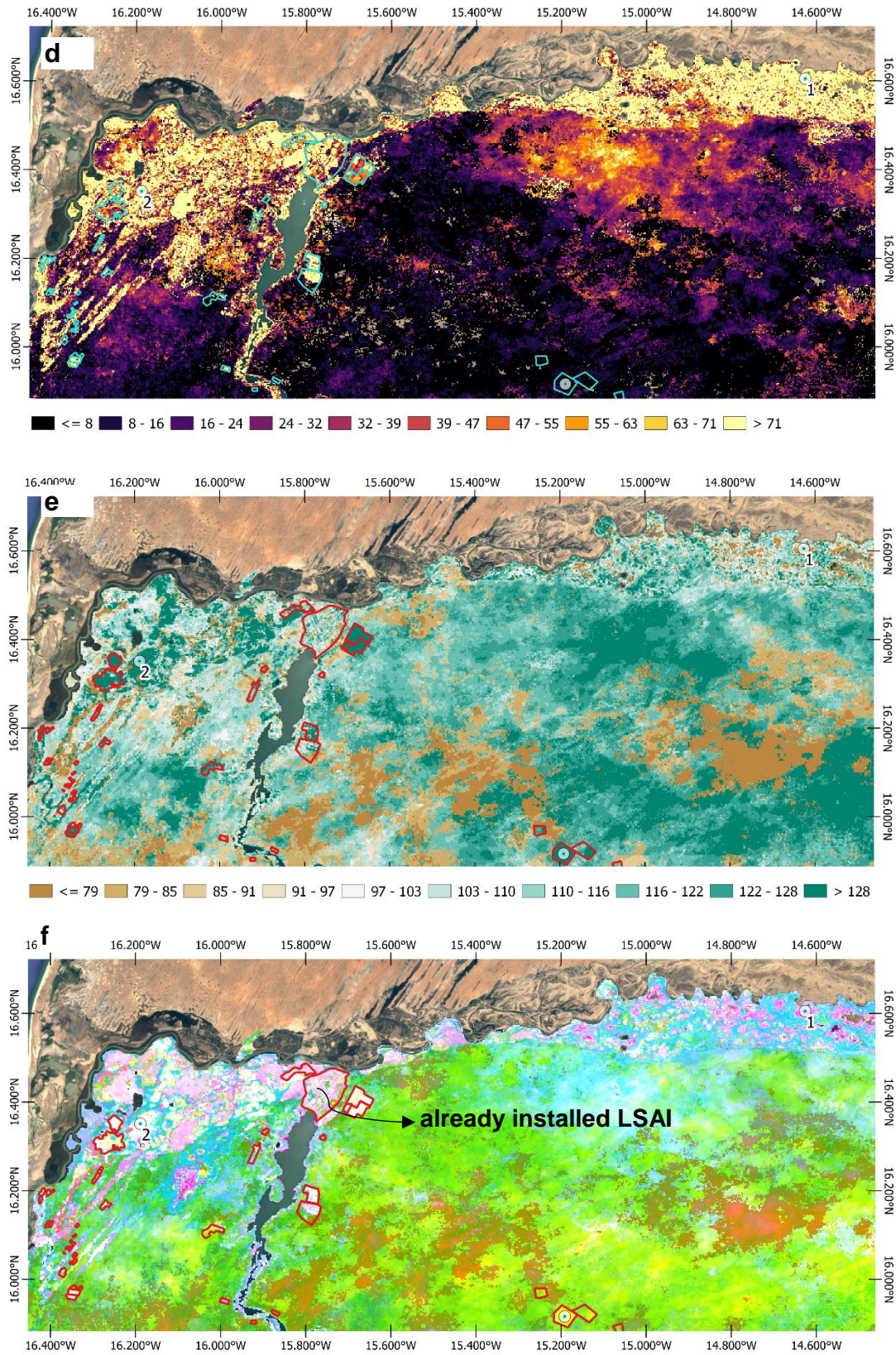


Appendix 3.B





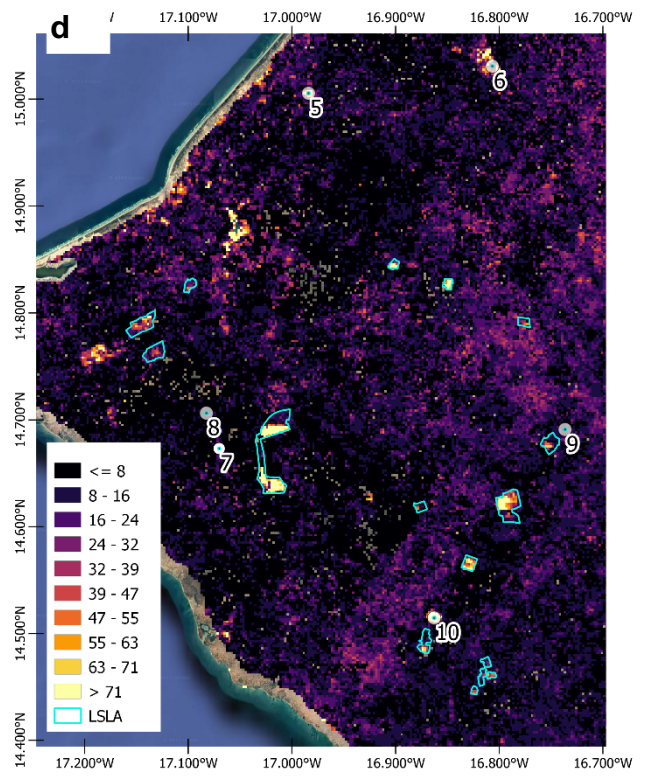
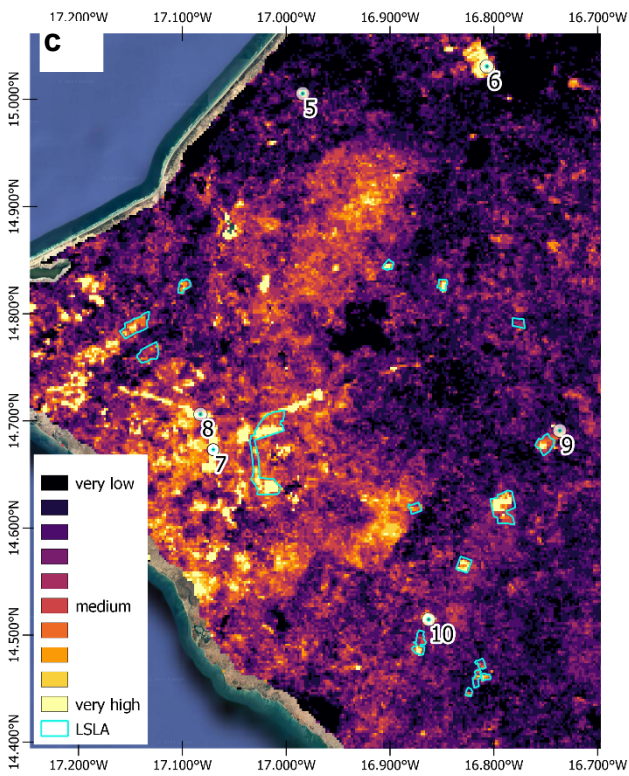
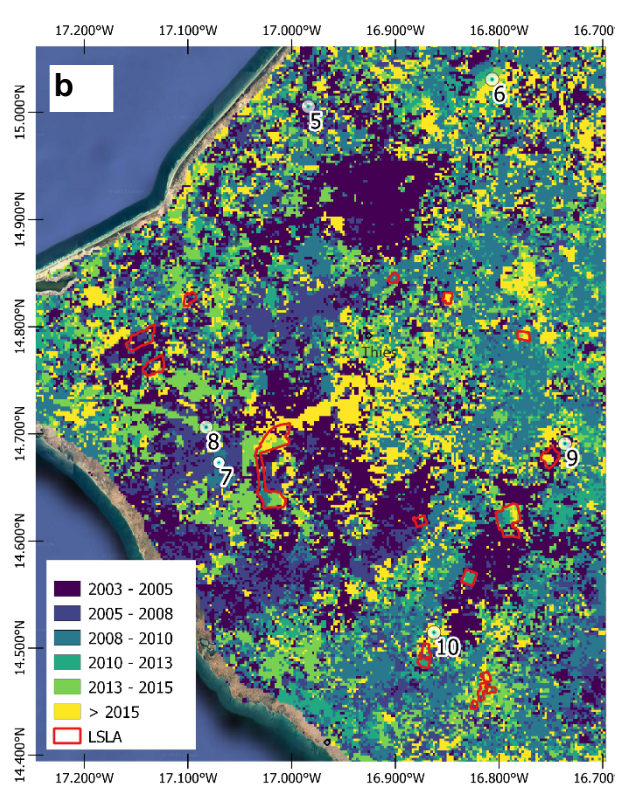
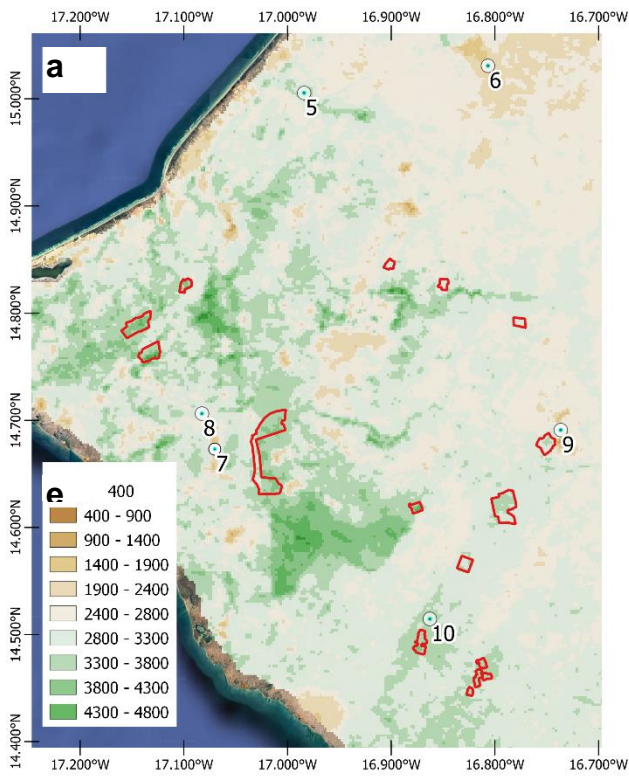
Chapter 3 : Appendices



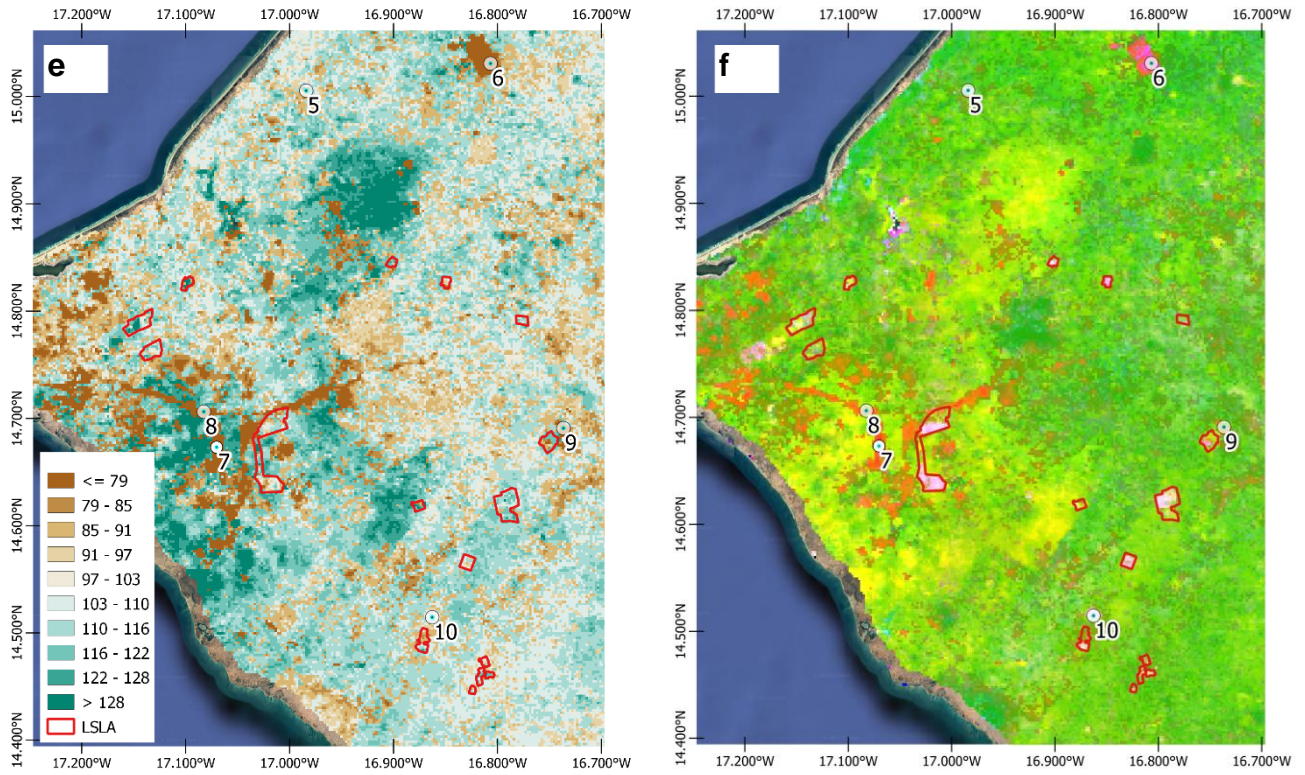
Close-in views of the Senegal River. a) MODIS NDVI 2000-2021 average map, b) Date of change map, c) Magnitude of change map, d) Dissimilarity map, e) NDVI ratio map, f) RGB map



Appendix 3.C



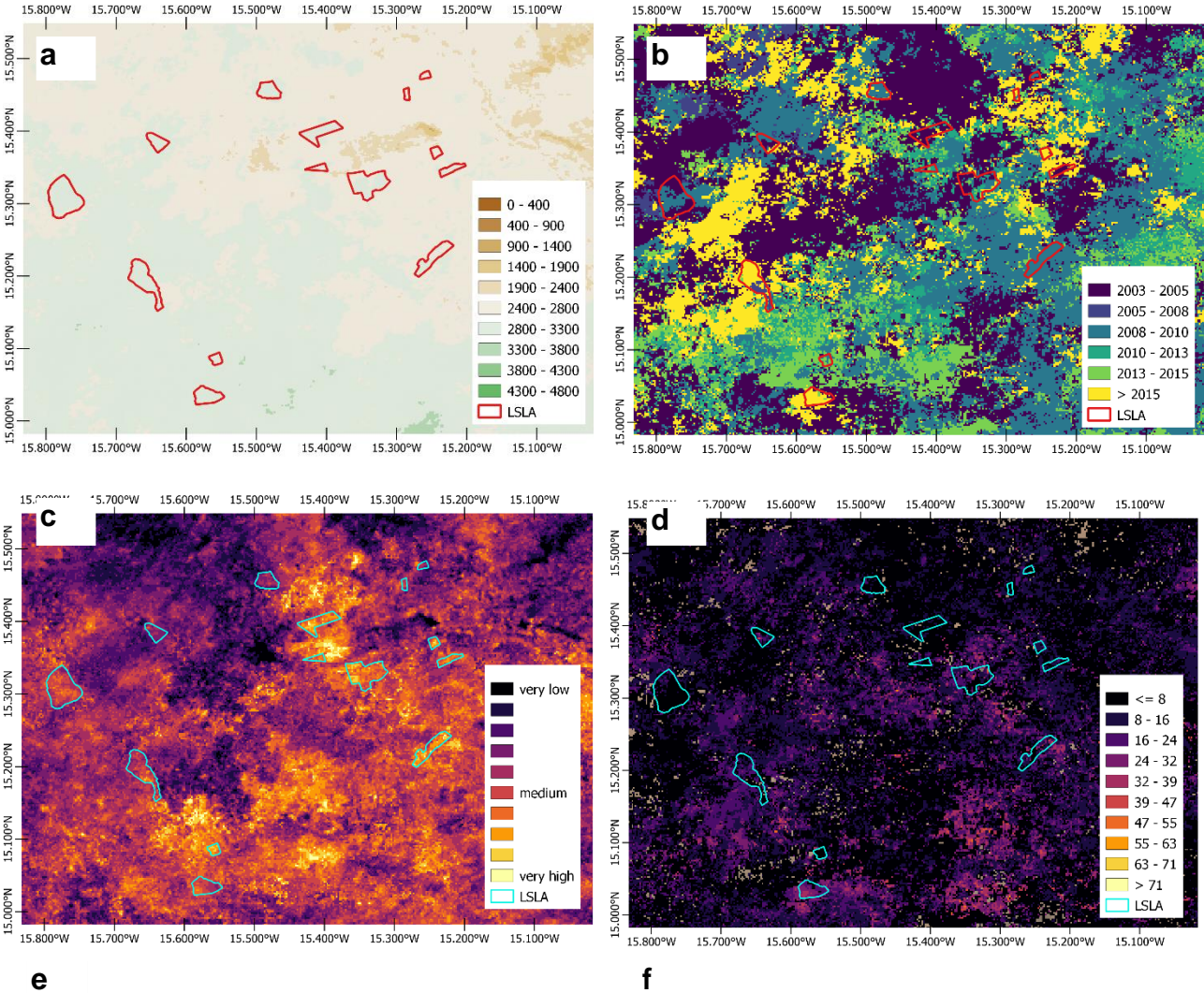
### Chapter 3 : Appendices



Close-in views of the Niayes. a) a) MODIS NDVI 2000-2021 average map, b) Date of change map, c) Magnitude of change map, d) Dissimilarity map, e) NDVI ratio map, f) RGB map.

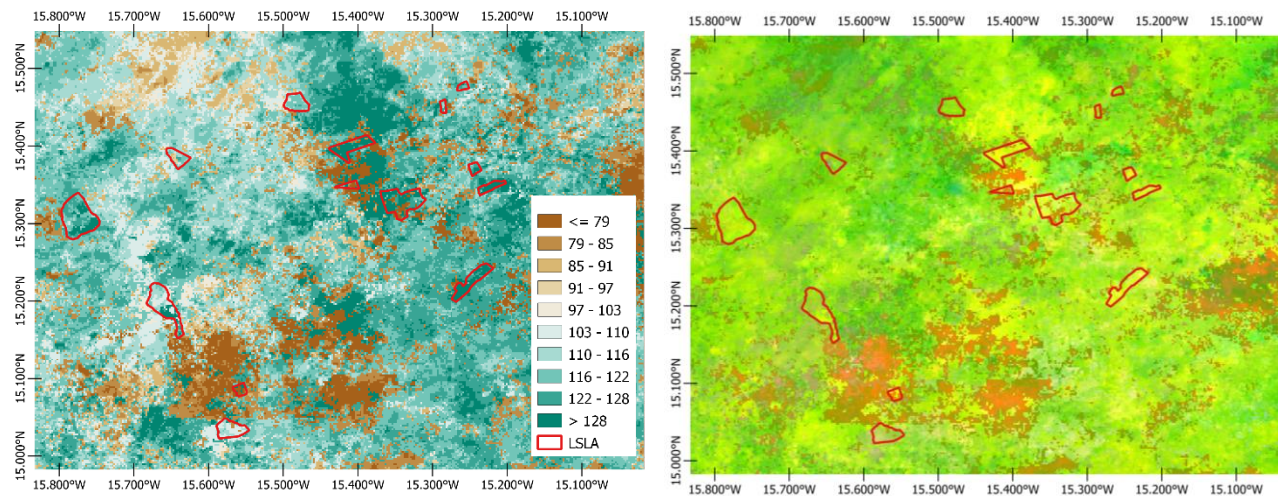


Appendix 3.D





### Chapter 3: Appendices



Close-in views of the Ferlo. a) MODIS NDVI 2000-2021 average map, b) Date of change map, c) Magnitude of change map, d) Dissimilarity map, e) NDVI ratio map, f) RGB map

# 4 MAPPING AND CHARACTERIZATION OF LARGE-SCALE AGRICULTURAL INVESTMENTS IN SENEGAL

---

This chapter is based on:

Ngadi Scarpetta, Y., Lebourgeois, V., Laques, A.-E., Dieye, M., Bégué, A.. Mapping and characterization of large-scale agricultural investments in Senegal. *In preparation*.

## 4.1 HIGHLIGHTS

- Dissimilarity metric improves the LSAs highlighting in the BFASTm-L2 change map
- LSAs share similar spectro-temporal characteristics per ecoregion
- 25-50% of field-reported LSAs are detectable in a fully unsupervised approach

## 4.2 ABSTRACT

The profound impact of human activities on the Earth's land surface over the past millennium has raised concerns about the sustainability of ecosystem services and human well-being, necessitating global monitoring of land use and land cover change (LULC). Among the major drivers of LULC change, agricultural expansion, often associated with deforestation, urbanization, and climate change, is a prominent contributor. **Large-scale agricultural investment (LSAI)** accounts for a significant proportion of this expansion. The rapid growth of LSAs over the past two decades, coupled with a lack of transparency, highlights the need for monitoring approaches to mitigate potential negative impacts.

To fill this gap, remote sensing offers a cost-effective and efficient solution for monitoring these complex land use systems over space and time. Various strategies have been used to detect agricultural LULC changes, with supervised classification methods being the most common. These methods, including pixel-based machine learning algorithms and object-based approaches, rely on data-driven approaches and the availability of representative training data, which is a major limitation for LSAs for which geographic data is scarce.

This study presents an unsupervised, remotely sensed approach to identify distinctive change characteristics of LSAs in Senegal, distinguishing them from other major drivers of change, such as natural (i.e. biotic and abiotic such wetlands) and anthropogenic drivers of biomass removal (e.g. urbanization, infrastructure development, and mining). The methodology involves a three-step process applied independently to the Niayes and Senegal River (SR) regions, known for their high LSAI presence. First, indiscriminate LULC changes were detected at the national scale using the fast and unsupervised BFASTm-L2 algorithm applied to MODIS NDVI SITS. The magnitude of change was then weighted using two change metrics: the time series shape dissimilarity and the NDVI ratio pre/post change to highlight potential LSAs. Second, data-driven contour analysis was used to extract **hotspots of change**. By cross-referencing these hotspots with Google Earth imagery and field data on Large-Scale Land Acquisition (LSLA), we aimed to identify the primary drivers of LULC change and build a validation database. Third, for each hotspot, object-based spectro-temporal features (i.e., the above-mentioned change metrics computed from MODIS and the NDBal index computed from

Landsat), as well as Landsat-based textural (Haralick features) and structural features (related to detected lines and rectangular/ circle shapes) were derived. Their effectiveness in discriminating LSAs was investigated using unsupervised methods. The assessment was performed using an LSAI field database.

The results show that the extracted hotspots of change, obtained by applying BFASTm-L2 and contour analysis of the weighted magnitude of change map, overlapped with 53% (in the Niayes) and 24% (in the SR) of the LSAs reported in the field database. Furthermore, unsupervised analysis of the hotspots shows that LSAs in each region can be distinguished from other land dynamics by their spectro-temporal and structural characteristics. Due to the varying discriminative power of structural features depending on the region analyzed, they were excluded from the unsupervised classification analysis to maintain methodological robustness and genericity. Results of a K-means clustering analysis based solely on the spectro-temporal features show that the precision rates of the most related LSAI clusters to LSAs are of 65% (in the Niayes) and 75% (in the SR).

While mostly exploratory, this research contributes to the development of robust and interpretable change detection methods that minimize reliance on external data, enabling the detection and comprehension of LSAs in regions with limited geographic information. These findings have implications for improving monitoring and understanding of LSAI dynamics in the context of global LULC change.

### 4.3 INTRODUCTION

Over the last thousand years, approximately 75% of the planet's land surface has been altered by human activities, threatening the sustainability of ecosystem services and human welfare, and urging the need for global monitoring of land cover and land use change (LULCC) (Radwan et al., 2021; Winkler et al., 2021).

Agricultural expansion, along with urbanization, deforestation (often closely linked to agricultural expansion) and climate change, is one of the main drivers of LULC change worldwide. It involves the conversion of land use from often natural environments to cultivated fields and can take many different forms, depending on factors such as location, scale, purpose, and environmental conditions. Amongst these, (active) **Large-Scale Agricultural Investment (LSAI)**, often driven by commercial enterprises and agribusinesses, would account for 23.8 million ha in 2016, of which 42% would be in Africa, representing approximately 10 million ha (37% of global reported area) (Nolte et al., 2016). Because of the speed and scale at which these large-scale investments are expanding, and the current lack of transparency surrounding most of these deals, monitoring their spatio-temporal dynamics is essential to mitigate potential negative impacts. This is all the more important given the

significant environmental, economic and social impacts of these structures, as well as the impact they may have on multiple issues such as food security (Davis et al., 2014; D'Odorico et al., 2017).

To fill this gap, remote sensing appears to be an appropriate, cost- and time-efficient tool to monitor these particular land use systems in space and time. Since LSAs are complex land use systems with many different forms (in terms of crops, environment, practices...), they cannot usually be derived directly from remotely sensed imagery. To overcome this, the most common approaches to date have been methods that focus on a specific region or cropping system, and take advantage of the particular spatial arrangement and temporal patterns driven by LSAs to derive spatio-temporal features that can be used as proxies for land use intensity. A commonly used proxy is cropping intensity, usually defined as the number of peaks per year (i.e., cropping cycles) in a pixel's vegetation index time series (Hentze et al., 2017). Another proxy for land use intensification is related to the availability of irrigation. Since irrigated agricultural land is expected to be more productive, temporal differences in biomass and greenness indices are exploited, especially during the dry season (Eckert et al., 2017). A third common proxy for land use intensity is related to field size and field texture, which are often larger and more homogeneous in LSAs due to mechanization (Graesser et al., 2018; Kuemmerle et al., 2009).

Methods to derive these proxies include very popular pixel-based machine learning algorithms (e.g., random forest) (Bey et al., 2020; Chen et al., 2023a), but also object-based approaches (Vogels et al., 2019), and deep learning-based methods that focus on detecting specific structural elements in the landscape, such as irrigation pivots (Tang et al., 2021). In most cases, these different strategies depend heavily on the availability of reliable and representative training data, which can be very difficult to obtain for LSAs, especially in remote regions such as sub-Saharan Africa (Bourgoin et al., 2019). Furthermore, these data, when available, are often outdated and incomplete due to the high spatio-temporal dynamics of LSAs. This severely limits the representability of the training data sets and thus the generalization ability of the supervised models. In addition many of these approaches make use of ancillary data that may not always be available or updated. For example, cropping masks are often used to limit the analysis complexity to cropland. However, these may be third-party products that are often outdated and not available in other regions (Hentze et al., 2017), may be based on assumptions that are not applicable to different regions (Ajadi et al., 2021), or, as mentioned above, may rely on training data (Graesser and Ramankutty, 2017), that have their own limitations in terms of quality and representativeness. Finally, many of the approaches, especially the object-based ones, often require high to very high resolution satellite imagery, which hinders their application on a large scale due to computational burden

and imagery cost. Therefore, fast, robust change detection methods that rely minimally on external data are desired to avoid the development of multiple region- and crop-specific models.

In a previous study, we proposed BFASTm-L2, a time-series change detection approach sensitive to seasonal changes, to detect anthropogenic LULC changes such as those induced by LSAs (Ngadi Scarpetta et al., 2023). The results confirmed the hypothesis that newly implemented LSAs in a small area of Senegal modified land cover, which in turn induced detectable seasonal changes in vegetation index (e.g. NDVI) time series. Because the algorithm uses coarse resolution MODIS data, it was shown to be fast and easily upscaleable. It also has the advantage of not relying on ancillary data or spatial masking. However, since the detected changes are not specific to LSAs and may include other anthropogenic changes (e.g. mining) or (a)biotic changes, a second study aimed to gain insight into the relationship between the main drivers of change and the induced types of change (i.e. abrupt, gradual, seasonal: amplitude, length of season (LOS), number of seasons (NOS)) and to derive some change metrics that, when combined, allow a better differentiation between them (Ngadi Scarpetta et al., 2024).

This study, which follows on from the two mentioned above, has two main objectives. The first is to **automatically locate and extract potential hotspots of change related to LSAs** based on BFASTm-L2 applied to MODIS 2000-2021 NDVI SITS (i.e., monitoring period between 2003-2018) and two MODIS-based derived change metrics: time series dissimilarity and NDVI post/pre change ratio (Ngadi Scarpetta et al., 2024). The second objective is to explore, based on an unsupervised approach, some common change characteristics of LSAs in Senegal that distinguish them from other major drivers of change, such as natural changes (biotic and abiotic changes) and anthropogenic drivers of biomass removal (e.g., urbanization, infrastructure development, and mining). These features should be as few and general as possible. As a result, remote sensing features that might be too specific to certain crop types or require complex settings to extract were avoided. This was done to better understand what characterizes LSAs, and with the idea of developing a fully automated pipeline in the future. The features include object based spectral features (i.e. MODIS derived change metrics and Landsat derived NDBaI) as well as Landsat based textural (i.e. Haralick features) and structural features. The next section describes the approach in detail.

## 4.4 DATA AND METHODS

### 4.4.1 Approach

#### 4.4.1.1 Workflow

The methodology consists of a 3-step approach, from extracting potential LSAI-related hotspots of change to distinguishing them from other land dynamics (Figure 4.1):

1. In the first step, indiscriminate LULC changes were detected at the national scale using the fast and unsupervised BFASTm-L2 algorithm (Ngadi Scarpetta et al., 2023) applied to long and dense MODIS NDVI time series. The magnitude of change was then weighted using the time series shape dissimilarity and NDVI ratio presented in Ngadi Scarpetta et al. (2024) (submitted for publication) to give more weight to potential LSAIs in the resulting change map.
2. From the weighted change map, hotspots of LULC changes were automatically extracted using a contour analysis, and labelled according to one of the main LULC change drivers identified using Google Earth and LSAI field database: LSAI, SA (smallholder agriculture), WET (wetlands), NAT (changes in natural environments other than wetlands), MIN (mining), INFR (infrastructure development), URB (urbanisation), MIX (mixed class) and FLOOD (floodplains).
3. The extracted objects of change (i.e. hotspots) were then characterized using spectro-temporal, textural and structural features derived from MODIS and Landsat, and clustered using the K-means algorithm based on the most discriminating and easily computable (in terms of genericity, i.e. without strong parameterization) object-based features. Clusters more strongly associated with LSAIs were identified and characterized.

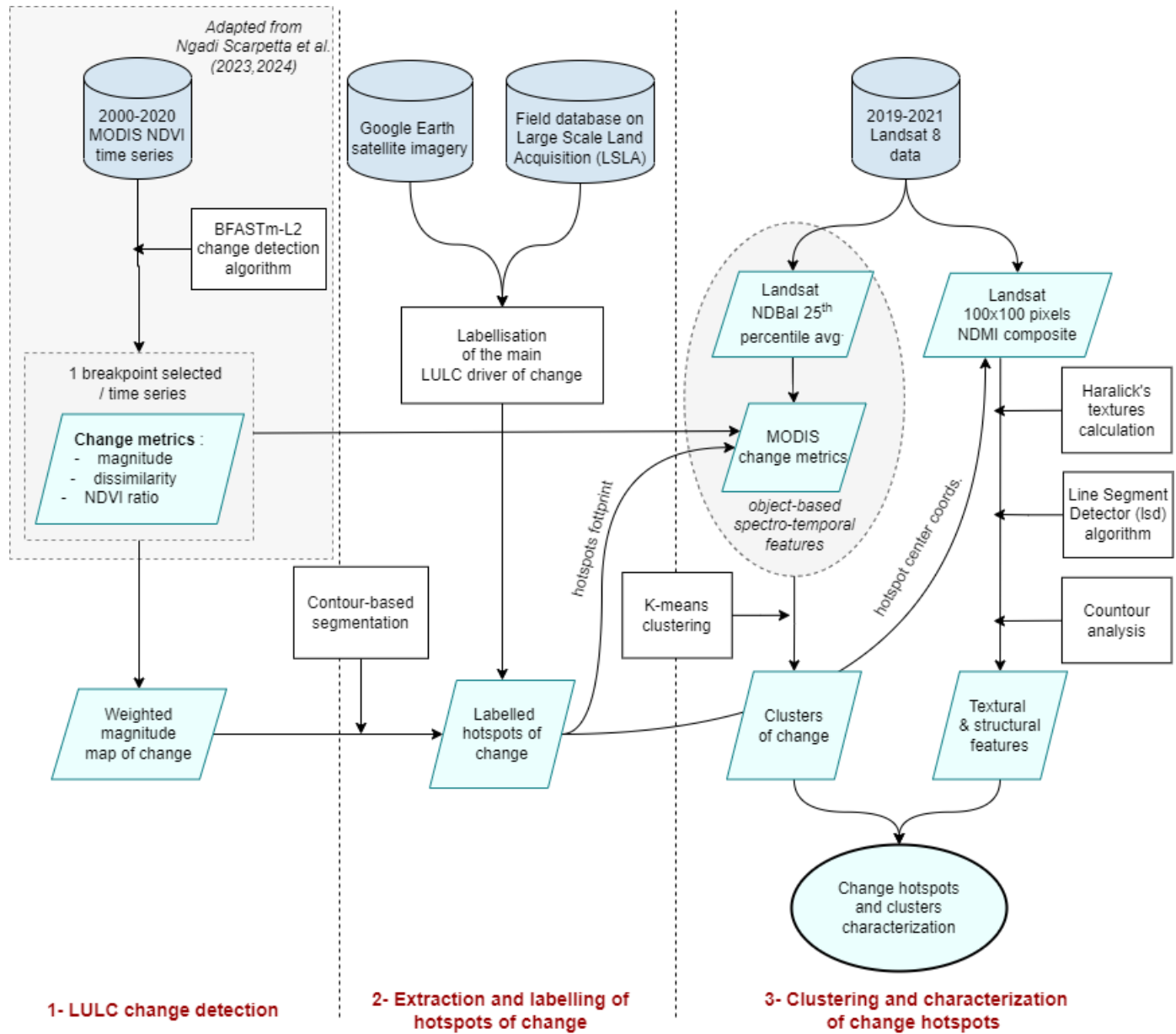


Figure 4.1: Three-step research workflow

#### 4.4.1.2 Study area

Located in the westernmost region of the Sahel, Senegal has a marked variation in rainfall from north to south, resulting in a semi-arid climate in the northern part (receiving 200-400 mm of rainfall per year) and a tropical climate in the southern region (receiving 800-1200 mm per year). The country has two distinct climatic seasons: a dry season from November to May and a rainy season from June to October. The predominant land cover types include steppe, savanna, and sub-humid dry forests, as documented by Budde et al. (2004), Sultan and Janicot (2003; Tappan et al.) and Tappan et al. (2004). The country has unique ecosystems, such as the fertile wetlands along the Senegal River, which have been vital to small-scale farmers, herders, fishermen, and traders for centuries. Unfortunately, these ecosystems are



increasingly threatened by the construction of dams and the implementation of irrigated rice projects, as highlighted by Horowitz and Salem-Murdock (1993) and Tappan et al. (2004).

Agriculture, which contributes to 15% of the country's GDP, is mainly characterized by small farms of less than 5 hectares, according to Bourgoin et al. (2019). Notably, there is an increasing trend of large-scale land acquisitions (LSLAs), with foreign investors deals counting for more than 3% of Senegal's arable land in 2016, as reported by Harding et al. (2016). Due to the lack of transparency, limited geospatial information, and potential socio-environmental impacts, initiatives have been launched to inventory and map LSLAs, as described in the works of Bourgoin et al. (2019) and Nolte et al. (2016). Nevertheless, challenges remain due to the spatio-temporal dynamic nature of LSLAs in terms of location and time, coupled with differences in methodologies. To address these issues, there is an urgent need for automated and rapid approaches to effectively monitor the entire national territory.

Agricultural LSAs, hereafter referred to as Large Scale Agricultural Investments or LSAs, are mainly concentrated in two contrasting regions (Figure 4.2):

- 1- The arid northern Senegal River region (hereafter referred to as SR, upper subplot in Figure 4.2) is an important agricultural region with a growing number of LSAs, mainly focused on horticulture, sugarcane production and cereals, mainly rice. Wetlands in this region are typically marsh type, saturated with water for a significant part of the year, which distinguishes them from floodplains, which are periodically flooded and traditionally used for flood-recession agriculture.
- 2- The Niayes (lower subplot,) contains many LSAs dedicated to horticulture. The vegetation consists mostly of open agricultural parkland. In this study, an extended part of the Niayes was considered, with a part of the Western Peanut Basin, which includes higher tidal wetland vegetation of the estuarine and mangrove type, mainly located in the Saloum Delta.

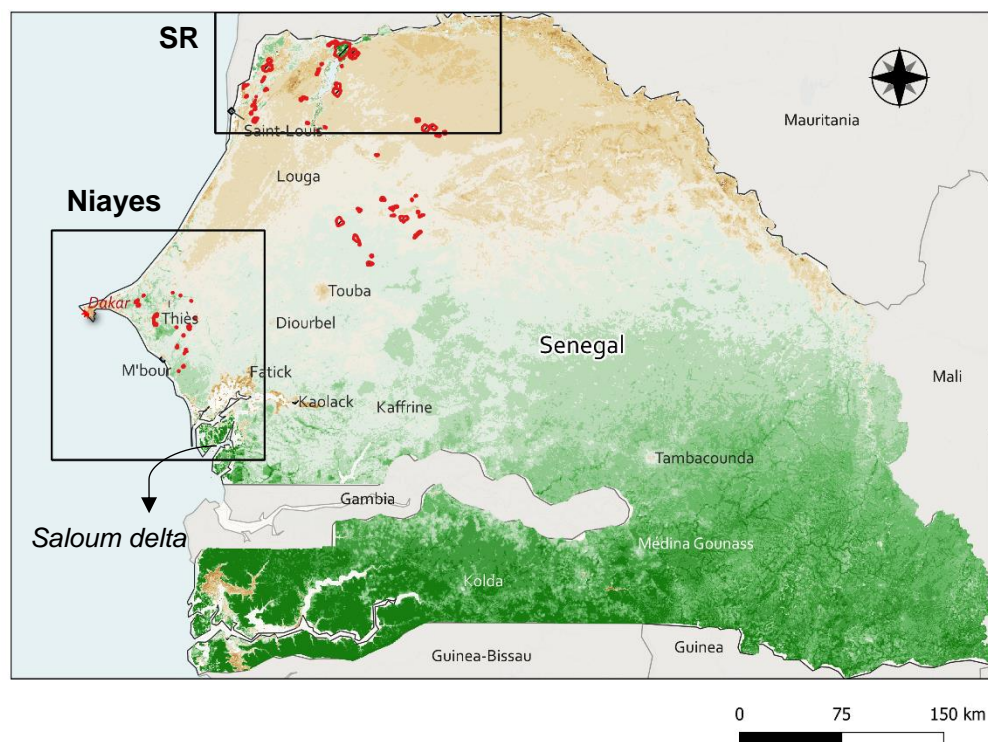


Figure 4.2: Overview of the study areas (black boxes): SR for the Senegal River region and an extended Niayes region. Red polygons represent the reported large scale agricultural investments within the field database. The base image represents the MODIS NDVI averaged over 2000-2020, with colors ranging from dark brown (lowest values) to dark green (highest values).

## 4.4.2 Data

### 4.4.2.1 LSAI field database

In 2019, the Senegalese Institute of Agricultural Research (ISRA) conducted a field campaign on **Large-Scale Land Acquisitions (LSLAs)** and recorded more than 700 polygons in a spatial database (M. Dieye, personal communication, 2022). The database contains deal information such as deal type (agribusiness, mining, etc.), size, year of transaction/negotiation. A filtered database was used, composed of 75 polygons containing only: i) agribusinesses created or expanded during the monitoring period of this study (2003-2018), ii) with a minimum size of 30 hectares, iv) with at least 1/3 productive area (assessed using Google Earth imagery). It is important to note that in this database some single LSLAs may be divided into many neighbouring blocks or digitised polygons. The distributions of LSAI size for the two studied regions are shown in **Figure 4.1** (see Appendix 4.E for more details).

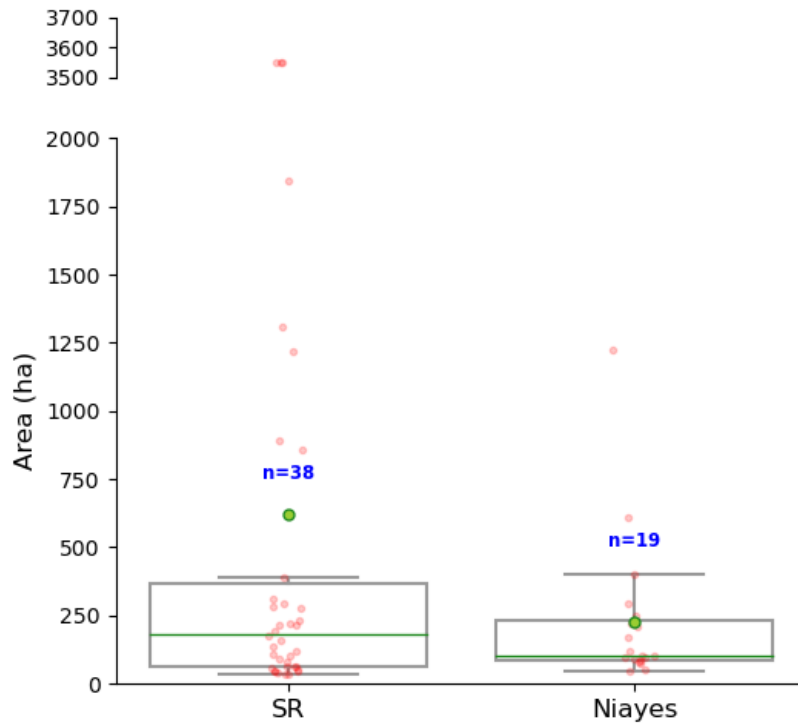


Figure 4.3: Boxplots of LSAI size (in hectares) by study region (medians are 181.5 and 104 ha for SR and Niayes respectively, indicated by the green horizontal line) and number of observations (n) in the ISRA field database. Green dots represent means.

#### 4.4.2.2 MODIS

The Moderate Resolution Imaging Spectroradiometer (MODIS) satellite sensor, in orbit since 1999, has been designed to improve our understanding of global dynamics and processes on Earth. With its global coverage, moderate spatial resolution (250 m) and high temporal resolution (1-2 days), this sensor is ideal for detecting subtle changes in land cover. Here, a set of MODIS NDVI 16-day composites at 250 m resolution (MOD13Q1, Collection 6) was acquired for Senegal over the period 2000-2021 and pre-processed in Google Earth Engine, using the same pre-processing as in Ngadi Scarpetta et al. (2023).

#### 4.4.2.3 LANDSAT

For each of the detected hotspots of change (step 1 of the workflow, see **Figure 4.1**) in each study region, Landsat 2019-2021 scenes with the highest available data quality (Landsat 8 Collection 2 Tier 1 and Real-Time data calibrated top-of-atmosphere (TOA)) were acquired using GE Engine. These data were selected because they follow the BFASTm-L2 monitoring period, which ends in 2019, and ensure land characterization following the detected change. Data have a spatial resolution of 30 m, sufficient for image analysis (i.e. textural and structural analysis) of LSAIs objects. Higher spatial resolution satellite imagery (e.g. Sentinel) was not considered here, as only the general structure of the LSAI (as structured by plot disposition) was of interest, while avoiding intra-plot variability. The data were first filtered for clouds and cloud shadows using the Quality Assessment (QA) band. Bits 3 and 5 (for cloud shadow and cloud respectively) were set to zero.

Data were extracted twice for each hotspot of change: 1) clipped to the hotspot footprint, and 2) a 100 x 100 pixel square image (approximately 900 hectares) centred on the hotspot centroid. These square images were considered necessary because some detected changes may represent only a small area of the LSAI (a few fields), from which it may be difficult to select the best contrasted composite image allowing extraction of structural and textural features (see section 4.4.3.4).

### 4.4.3 Methods

#### 4.4.3.1 LULC change detection

##### 4.4.3.1.1 The BFASTm-L2 weighted magnitude of change

In the first step of the approach, the BFASTm-L2 change detection algorithm (Ngadi Scarpetta et al., 2023) was applied to the 2000-2021 MODIS NDVI time series. The magnitude of change calculated by BFASTm-L2 corresponds to the most significant breakpoint detected in each time series, measured as the Euclidean distance between the two 3-year time series flanking the breakpoint. Consequently, only changes in the period 2003-2018 can be detected. However, as this distance is sensitive to trends and amplitude changes, it may capture natural and climatic variability-induced changes. To disentangle these contributions, and following Ngadi Scarpetta et al. (2024), two additional change metrics were computed: a time series shape similarity metric (hereafter **referred to as dissimilarity**) based on the Procrustes distance, and an NDVI post/pre change ratio based on 3-year averages (hereafter **referred to as NDVI ratio**). Procrustes analysis is a statistical shape analysis that optimally superimposes (by translation, rotation, reflection and scaling) two (or more) vectors (Gower, 1975). Values  $\in [0,1]$ , with higher values indicating higher dissimilarity. The function Procrustes of the Python scipy spatial package was used.

These metrics contribute to the F factor presented here below and increase the importance of shape changes in the overall magnitude of change calculation.

$$F = \left| \ln \left( \frac{\tilde{ts}_2}{\tilde{ts}_1} \right) \right| + \text{procrustes}(ts_{1_{agg}}, ts_{2_{agg}}) \quad \text{Equation 4-1}$$

With  $ts_1/ ts_2$  the 3-year subsample *before* / *after* the breakpoint; and  $ts_{(1(2))_{agg}}$  the same subsamples monthly averaged over the 3 years.

The first logarithmic term of F accounts for changes in amplitude. The logarithmic transformation allows most values to fall between [0,0.22] (in fact, it is rare for the NDVI post/pre-change ratio to exceed 1.25 in Senegal (Ngadi Scarpetta et al., 2024) ). The second term of F explicitly quantifies the shape similarity between  $ts_1$  and  $ts_2$ , and ranges from 0 to 1. This additive combination allows less weight to be given to very large abrupt changes that do not significantly alter the time series pattern. This factor was then incorporated into a weighted

magnitude of change to better emphasize LULC changes that are mostly seasonal in nature. The BFASTm-L2 weighted magnitude of change, used to derive the change map, was calculated as follows:

$$\textit{weighted magnitude} = \textit{change magnitude} \times F \quad \textit{Equation 4-2}$$

### 4.4.3.1.2 Assessment of the weighted magnitude of change to LSAs

The sensitivity of the weighted magnitude to LSAs was assessed by analyzing the difference in magnitude between the median values inside and outside each LSLA. Outside values correspond to the medians of the magnitudes within a positive 2 km buffer around each reported LSLA (reported in the field database). Differences are expressed in percent with respect to the outside values. Only the polygons with more than 66% of their area occupied (assessed with Google Earth imagery) were used for this analysis, resulting in 75 polygons located in two agro-climatic regions of Senegal. The idea behind this difference is to assess how contrasted the objects are with respect to their environment and how likely they are to be visually detected on a map with the weighted magnitude. The results are presented in the form of a stacked histogram, with two classes representing the two study areas (Senegal River valley and the Niayes).

### 4.4.3.2 Extraction and labelling of change hotspots

#### 4.4.3.2.1 Extraction of change hotspots

Hotspots of change were detected and extracted from the weighted magnitude of change map. The pre-processing of the weighted magnitude of change map included the application of a 500m buffer around the water (corresponding to pixels with negative MODIS 2000-2021 NDVI mean). The buffering allows to exclude highly unstable coastline pixels with very high magnitudes of change from the analysis. A Gaussian blur (kernel=7) was then applied to average rapid changes in pixel intensity before performing a linear stretch of the images between the 1<sup>st</sup> and 99<sup>th</sup> percentiles. Image segmentation was then performed using a 6-level contour segmentation. The Python *contour* function from the Matplotlib library was used. Aggregation of the last 2 levels was performed to obtain larger polygons. Then, the convex hull of each detected polygon was computed to obtain more "compact" shapes.

#### 4.4.3.2.2 Labeling change hotspots with a LULC class

To evaluate the discriminative power of the spectro-temporal, textural and structural features used in this study (and presented in the next section) for different drivers of change, each extracted hotspot of change (from the previous section) was labeled in one of the following LULC classes: LSAI, SA (smallholder agriculture), WET (wetlands), NAT (changes in natural environments other than wetlands), MIN (mining), INFR (infrastructure development), URB (urbanization), MIX (mixed class). In the northern Senegal River region, the WET class was further subdivided into WET (wetlands) and FLOOD (floodplains) classes due to its

specificities. The main drivers of change were visually assessed using the Google Earth imagery and the field database on LSAs.

### 4.4.3.3 Characterization of change hotspots

#### 3.6.1.1.1 Spectro-temporal features

As proposed by Ngadi Scarpetta et al. (2024) (submitted for publication), we used a combination of three MODIS-based (spectro-temporal) change metrics to highlight LSAs: the BFASTm-L2 magnitude of change, the time series shape dissimilarity, and the ratio of the 3-year NDVI average before and after the detected change. Because changes caused by LSAs may have some similarities to those caused by wetlands (Ngadi Scarpetta et al., 2024), and because we expect LSAs to have less frequently exposed bare soils, we introduced an additional spectral variable, the Normalized Difference Bareness Index (NDBal). This index is based on the difference between the strong emission of thermal infrared (TIR) radiation and the near total absorption of shortwave infrared (SWIR) wavelengths by bare soil (Zhao and Chen, 2005). It is effective in distinguishing bare ground from similarly constructed structures and vegetation. The NDBal was calculated as follows (following the Landsat band names)

$$NDBal = (SWIR1 - TIRS1) / (SWIR1 + TIRS1) \quad \text{Equation 4-3}$$

The Landsat 20019-2021 NDBal 25<sup>th</sup> percentile was calculated for each pixel, and then aggregated at the object-level (i.e. hotspot of change) using the mean.

The discriminative power of each spectro-temporal feature for LULC drivers of change was assessed through visual analysis of pairwise scatterplots and univariate distributions of the spectro-temporal features.

### 4.4.3.4 Textural and structural features

#### 4.4.3.4.1 Landsat NDMI composites

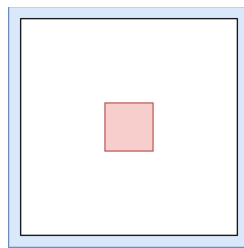
Once the boundaries of all detected change objects were obtained by the segmentation step, textural and structural features were computed for each object based on a composite image calculated to provide the best possible contrast between the target object and its surroundings. The 100x100 pixels composite, centered on the center of each detected change hotspot, was based on the NDMI (for Normalized Difference Moisture Index) vegetation index, used in studies aiming to detect irrigated agriculture (Chance et al., 2018) and computed as follows:

$$NDMI = (NIR - SWIR1) / (NIR + SWIR1) \quad \text{Equation 4-4}$$

After averaging the 2019-2021 NDMI images on a monthly basis (thus obtaining 12 monthly averaged images), a 3-month rolling average was applied to each monthly image. This is defined by calculating, at the pixel level, the NDMI average of the current month, the previous



month, and the following month. This allowed the values of a 3-month period (roughly equivalent to a growing season) to be integrated into each image, and allowed to obtain an image that best represents the growing season. **Then, the image with the highest contrast was then selected to better identify: (i) the growing season and (ii) the edges of the LSAs.** Contrast was calculated as the difference between the average pixel values in the center of the image (represented by the red square in Figure 4.4), and the average pixel values at the edge of the image (in blue). This requires that the LSAs are well separated from other land objects, as is the case in the Niayas. For places where this is not the case and the difference between the red and blue averages is less than 10 (a defined hard threshold), such as the agriculturally intensive SR region, the image with the highest interquartile range (IQR) (difference between the 75th and 25th percentile) was selected.



*Figure 4.4: Schematic view (to scale) of the 100 x 100 pixel Landsat image (full square). Coloured areas were used to identify the composite image with the highest contrast, representing the largest difference between the average NDVI value in the red area and the average NDVI value in the blue area.*

Because intensive agriculture, as a result of mechanization, has often been reported to produce more homogeneous landscapes with larger field sizes of regular shape, many studies have used textural features to distinguish LSAs from other land use systems (Bey et al., 2020; Kuemmerle et al., 2009). In this study, the fourteen textural metrics proposed by Haralick et al. (1973) were computed from the Gray Level Co-occurrence Matrix (GLCM), using the NDVI composite image clipped to each hotspot footprint as the base image. The *pyfeats* package (Giakoumoglou, 2021) was used, allowing to compute the following features: 1) Angular Second Moment (ASM), 2) Contrast, 3) Correlation, 4) Sum of squares (variance), 5) Inverse Difference Moment (or homogeneity), 6) Sum Average, 7) Sum Variance, 8) Sum Entropy, 9) Entropy, 10) Difference Variance, 11) Difference Entropy, 12-13) Information Measure of Correlation 1 and 2, and 14) Maximal Correlation Coefficient. As for the spectro-temporal features, the discriminative power of each textural feature for LULC drivers of change was assessed through visual analysis of pairwise scatterplots and univariate distributions.

#### 4.4.3.4.2 Structural features

The structural features were derived from the 100x100 pixel Landsat-based NDVI composite images. The goal was to derive features that characterize the shape geometry of the landscape objects present in the Landsat scene, assuming that the field geometry of intensive agriculture

is generally polygons with square angles and circles. Prior to feature extraction, the images were pre-processed, to improve image segmentation, which allowed the detection of object shapes and lines. This includes: 1- contrast stretching between the 2nd and 98th percentiles, 2- image quantization in 10 color bins, 3- bilateral smoothing, 4- image equalization, and 5- thresholding using the 7<sup>th</sup> computed Jenks natural break (Jenks, George F., Caspall, Fred C., 1971).

Then, two methods were tested for the detection of shapes and lines, keeping the same parameterization for both regions (SR and Niayes). The first was by image segmentation using a contour analysis (Python *contour* function from the Matplotlib library) applied over the whole 100x100 pixels image. The number of square angles (i.e.  $90^\circ \pm 10^\circ$ ) and circles was calculated, and correspond to the ***cnt counter*** metric. Only the polygons that: 1- were not too large (i.e. < 2000 pixels; very large objects often correspond to natural objects, while objects such as LSAs are expected to be segmented into smaller objects corresponding to fields); 2- had at least one square angle or a *circularity* index close to 1 (i.e. >0.8 ) (Cox, 1927) , 3- did not have a very complex shape, i.e. a *Shape index* close to 1 (i.e. < 1.6 ) (McGarigal, 2002) were considered.

In the second method, line detection was performed using the Line Segment Detector (*lsd*) algorithm (Grompone von Gioi et al., 2012), implemented in the Python OpenCv *LineSegmentDetector* function. Default parameters were used. Unlike the contour-based method, which uses the contours extracted within the entire image, this method was only applied within the hotspot footprints. This minimized the number of "false" detections, since all segmented objects (also outside the hotspot footprint) are composed of lines. The number of parallel lines and the number of square angles (intersection angle of  $90^\circ \pm 10^\circ$ ) were then calculated and standardized to the footprint area (***line counter***).

The two computed metrics, i.e. ***cnt counter*** and ***line counter*** were then combined into a new variable, called ***geom index***, corresponding to:

$$\mathbf{geom\ index} = \ln( (\mathbf{cnt\ counter} + 1) \times (\mathbf{line\ counter} + 1) ) \quad \text{Equation 4-5}$$

As with the previous features, the univariate distributions of the three calculated metrics were plotted for each type of LULC driver of change, allowing a visual assessment of the discriminative power of each metric.

#### 4.4.3.5 Clustering of change hotspots

The detected land objects of change, previously labelled into main LULC classes, were clustered using the K-means algorithm. This was done in the hope of detecting a group of similar objects that could be related to LSAs in an unsupervised manner and thus help in their characterization. Only the discriminant (according to the univariate distributions plotted for



each type of LULC driver of change) and generic (requiring few parameters in their computation) features were used in the clustering.

The best partitioning (i.e., K-means k parameter) was identified using the Silhouette metric, for partitions ranging from 2 to 10. The silhouette metric is a measure of how similar an object is to its own cluster (cohesion) compared to other clusters (separation). The silhouette values range from -1 to 1, with high values indicating a good fit, i.e. good cohesion within clusters and separation between clusters, and values near or below zero indicating a very poor fit and misclassification. In this study, and for similar overall Silhouette values, the partitioning that produced the smallest number of clusters (k) was chosen to improve the interpretation of the results, which consist of a box plot of the clusters against the features used in the classification.

### 4.4.3.6 Evaluation of the performance in detecting LSAs

The evaluation consisted of two steps:

1. The first step evaluated the performance of the BFASTm-L2 change detection algorithm in detecting LULC changes induced by LSAs. This was done by counting the number of extracted hotspots of change that overlap the LSLAs (i.e digitized polygons) reported in the field database.
2. The second step was to evaluate the precision of the clustering with respect to LSAs. In this step, the clusters related to LSAs were first combined into a unique group. Then the precision of this unique group to LSAs was calculated as follows:

$$\textit{precision} = \textit{True positives} / (\textit{True positives} + \textit{False positives}) \quad \textit{Equation 4-6}$$

The true positives correspond to the number of objects within the combined group that are LSAs. The false positives correspond to the number of all other objects.

## 4.5 RESULTS

### 4.5.1 LULC change detection

#### 4.5.1.1 Change maps

To improve the contrast of potential LSAs, we here proposed a change magnitude weighted by the dissimilarity and NDVI ratio as explained in section section 2.3.1. Maps of these weighted magnitudes are shown in **Figure 4.5** for each study region.

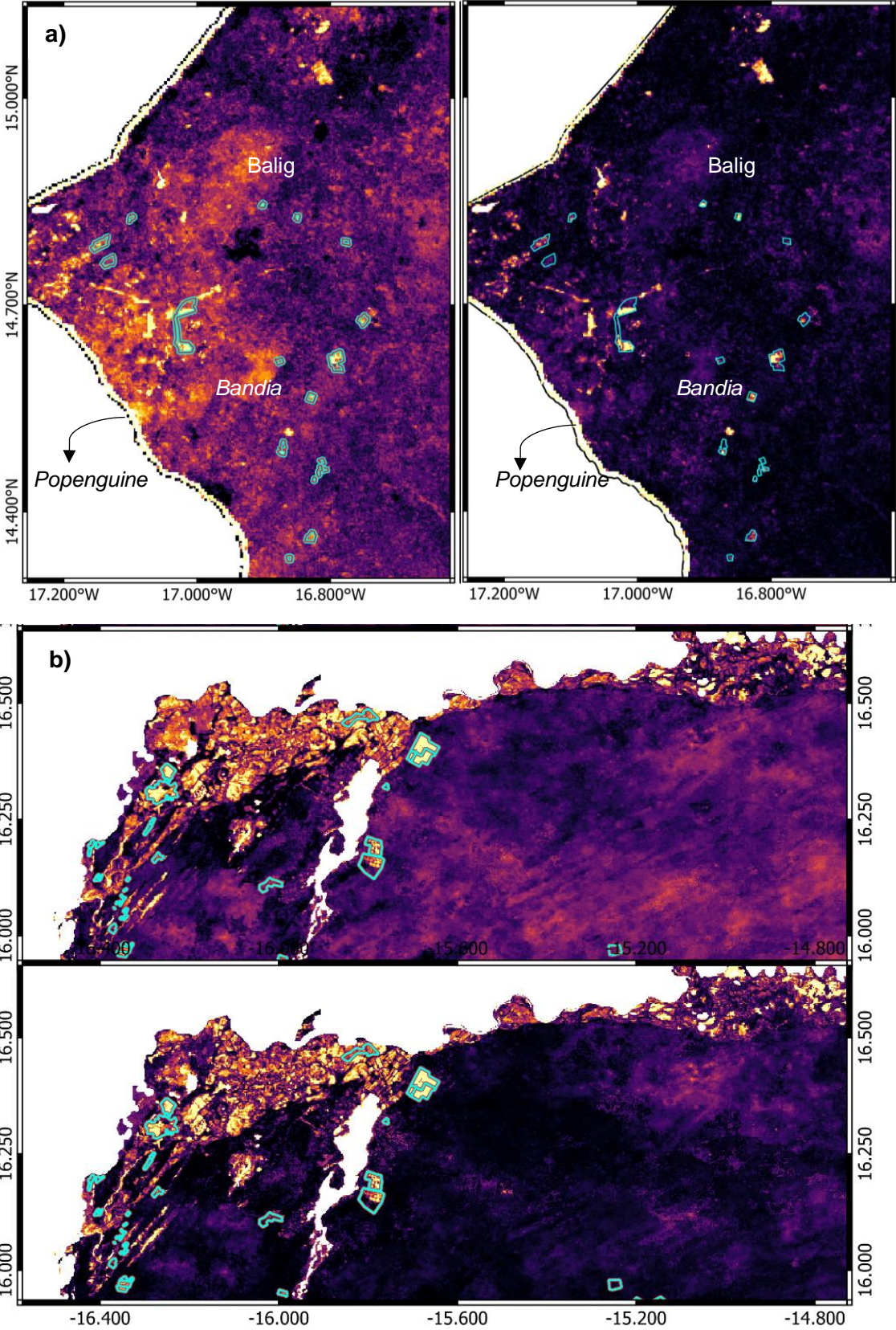


Figure 4.5: Change maps for the Niayes (a) and the Senegal River regions (b). From left to right (a)/ top to bottom (b): the BFASTm-L2 magnitude of change map, the magnitude of change map weighted with the dissimilarity and NDVI ratio metrics. LSAIs reported in the field database are shown in cyan.

Overall, background contributions are minimized for both regions using the weighted magnitude of change (subplots a vs. b) while it is less the case when the magnitude is not weighted. In the Niayes, diffuse areas of change occurring within forested areas (Bandia, Popenguine, area near Balig...) are removed, leaving clear contributions from LSLAs, mining and infrastructure activities characterized by large seasonal changes (in NOS/LOS) and/or large abrupt changes. In the arid region of the Senegal River, the contrast seems to be less important within the intensive agricultural area in the northwest, although a minimization of the background contributions is also observed.

To better assess the contrast enhancement observed in the maps when using the weighted magnitude of change, quantification of the contrast is performed in the next section using the LSLAs reported in the field database.

### *4.5.1.2 Assessment of the weighted magnitude of change to LSAIs*

The performance of BFASTm-L2 in detecting agricultural LSLAs was evaluated here by analyzing the difference in magnitude (weighted or not) between the median values inside and outside each reported LSLA (see Section 2.3.5). Differences are expressed in percent with respect to the outer values. This allows an assessment of how different objects are from their surroundings, and thus how likely they are to be automatically detected on a map.

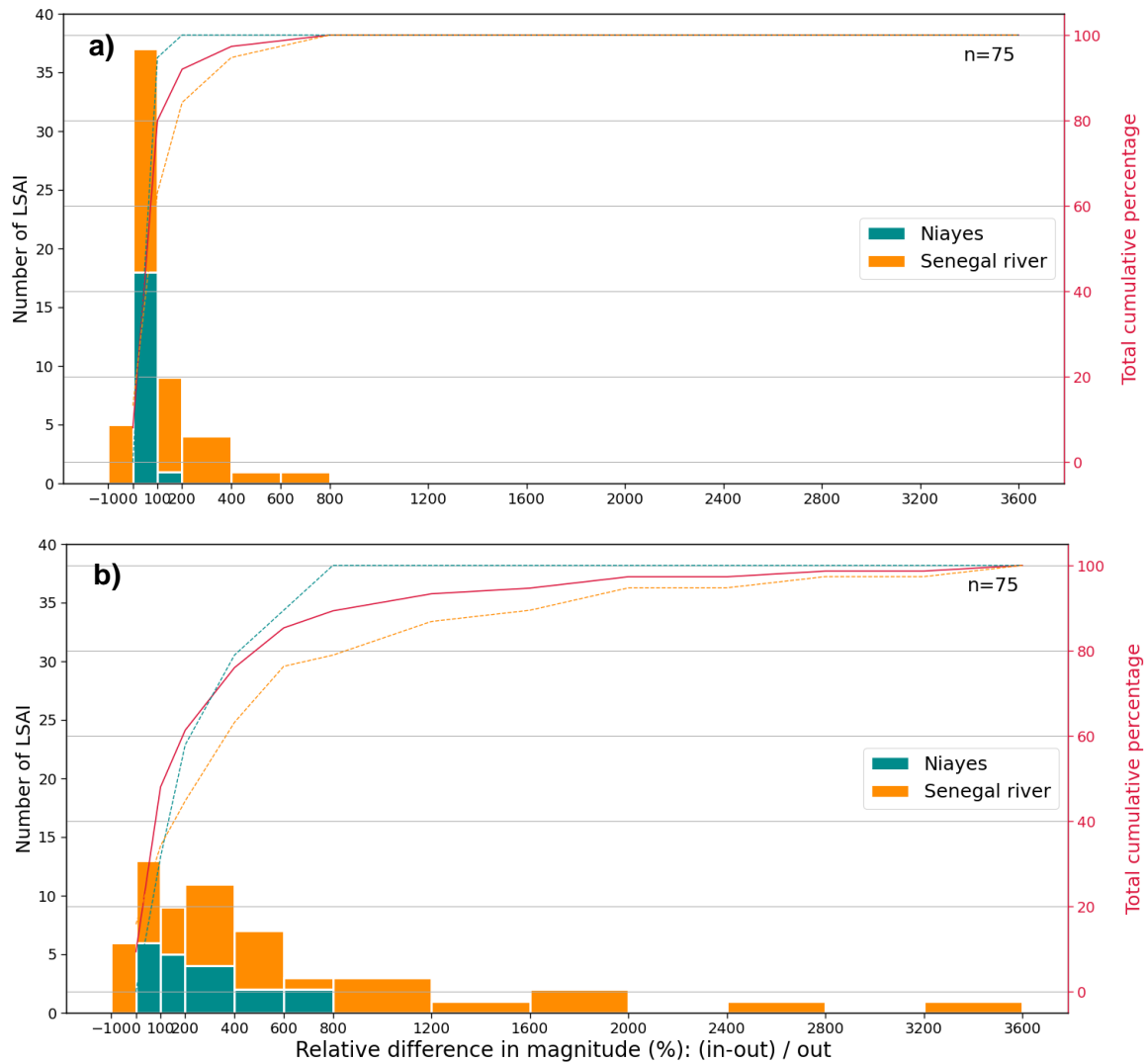


Figure 4.6 Histogram of the relative differences (in %) between the medians of the magnitudes of change within ("in") and outside ("out", i.e. within a 2 km buffer around the LSAI) the LSAIs in the field database. a) Using the BFASTm-L2 magnitude of change, b) Using the weighted magnitude of change.

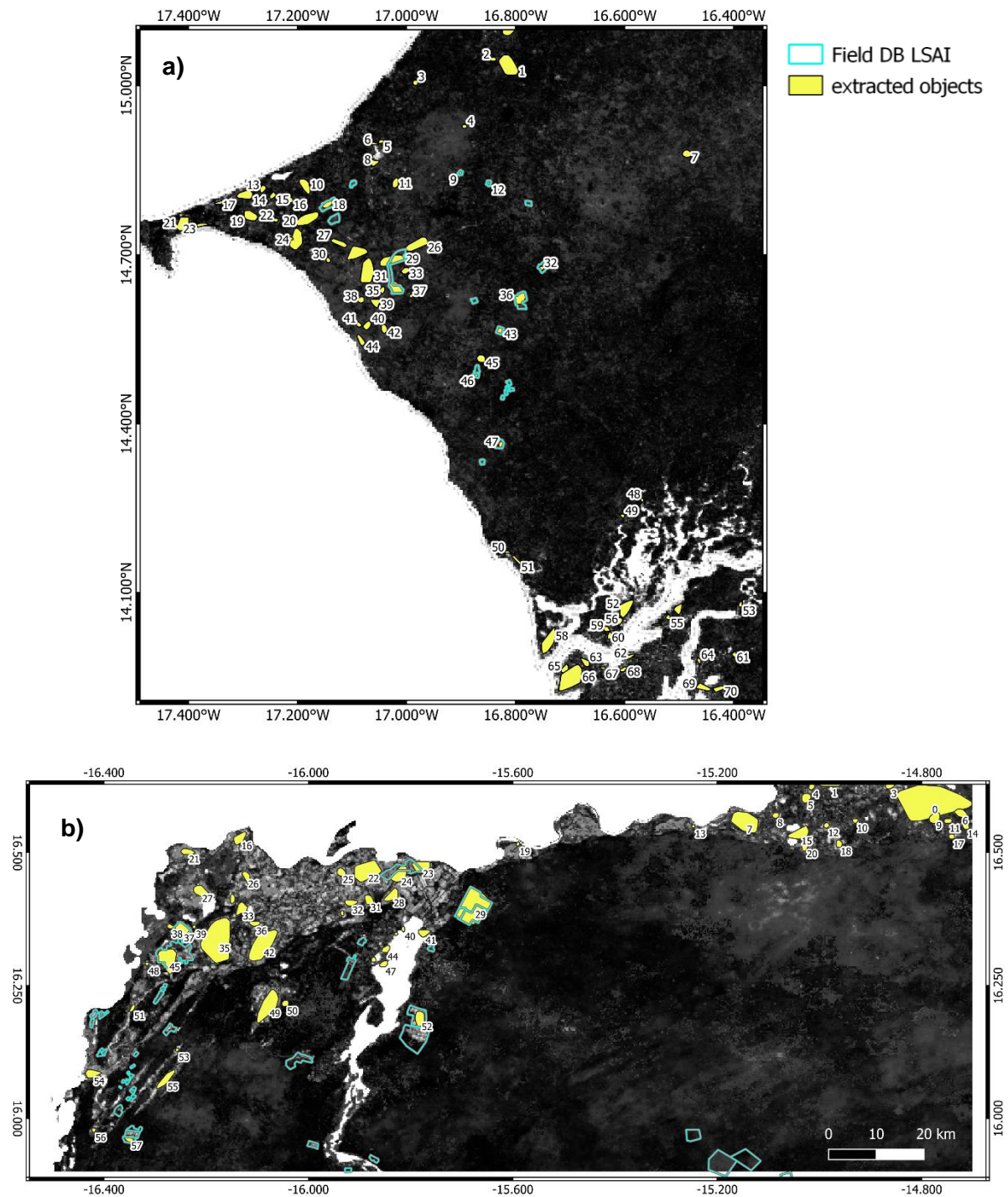
**Figure 4.6** shows the histograms of the differences using two different change metrics: the BFASTm-L2 magnitudes of change and the weighted magnitudes of change (subplot a and b, respectively). A first observation is that all distributions are more spread out when using the weighted magnitudes of change, indicating a better contrast between the LSLAs and their surroundings, with about 10% to 40% of the samples having a difference above 200% (see subplot a and b respectively). In the Niayes, 0% to ~40% of samples have a difference > 200%, and in Senegal River ~20% to ~60% of samples have a difference > 200%. These results show the ability of the weighted magnitude to detect and highlight changes caused by LSLAs.



## 4.5.2 Extraction and characterization of change hotspots

### 4.5.2.1 Extracted hotspots

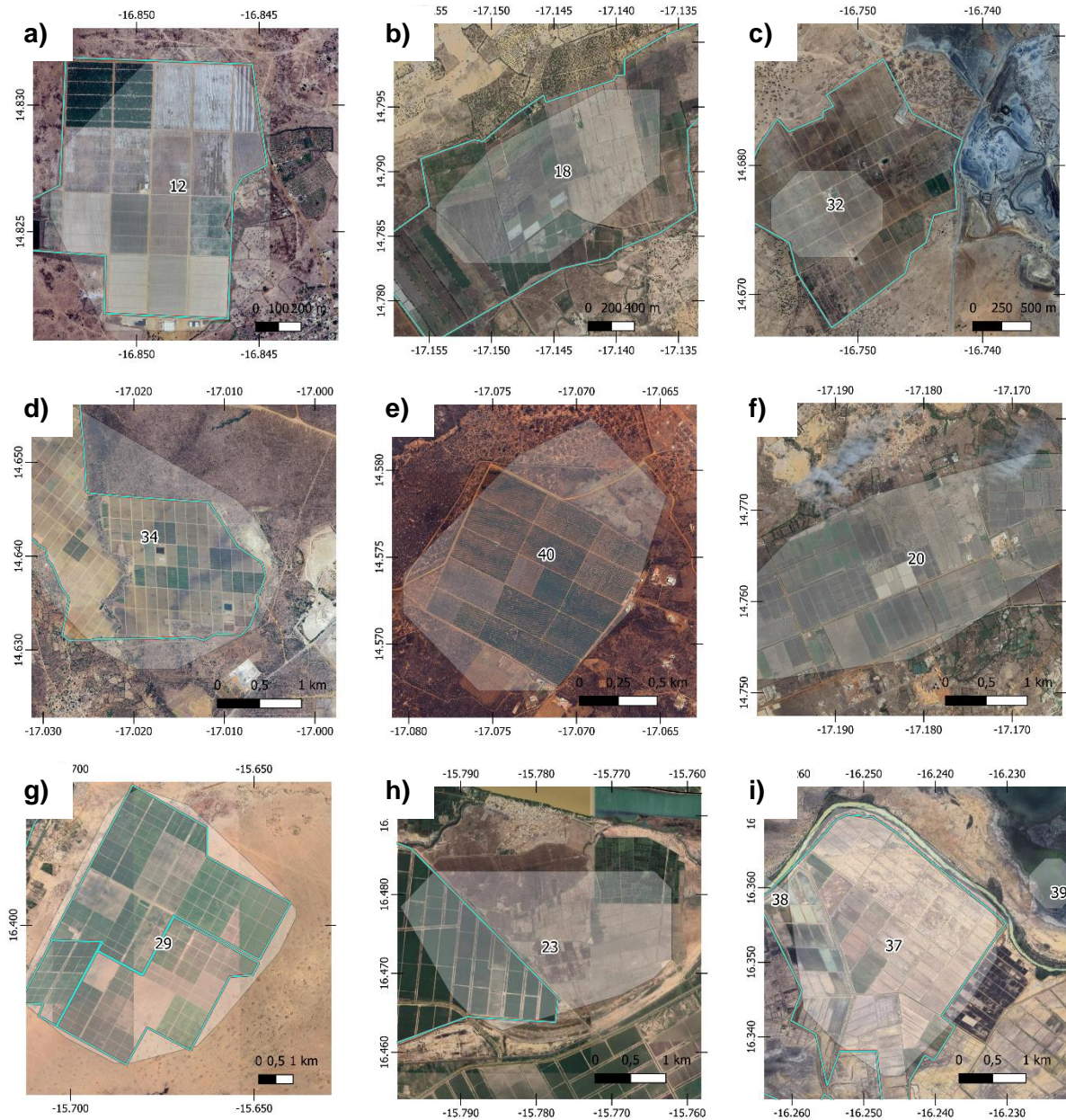
Hotspots of change extracted for the Niayes and the Senegal River regions following the approach detailed in Section 3.3.3 are presented here below, in **Figure 4.7 a and b respectively**.



## Chapter 4: Results

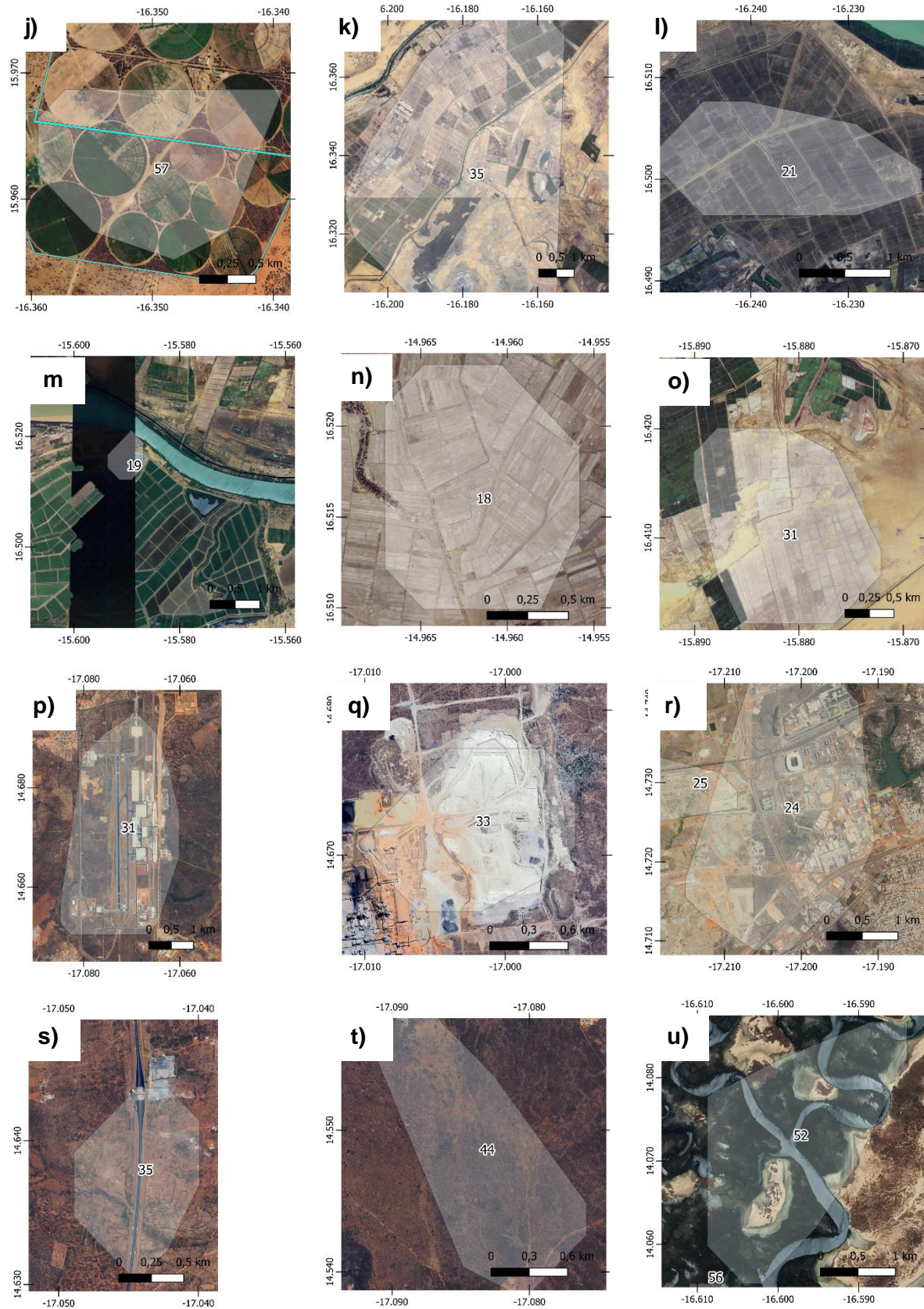
Figure 4.7: Study regions with LSAs from the non-exhaustive ISRA field database (in cyan) and extracted hotspots of change (yellow). a) The extended Niayes region; b) the Senegal River (SR) region. Base map correspond to the weighted magnitude of change.

A sample of the variety of landscape objects detected is shown in **Figure 4.8**, with high resolution imagery from Google Earth.





## Chapter 4: Results



## Chapter 4: Results

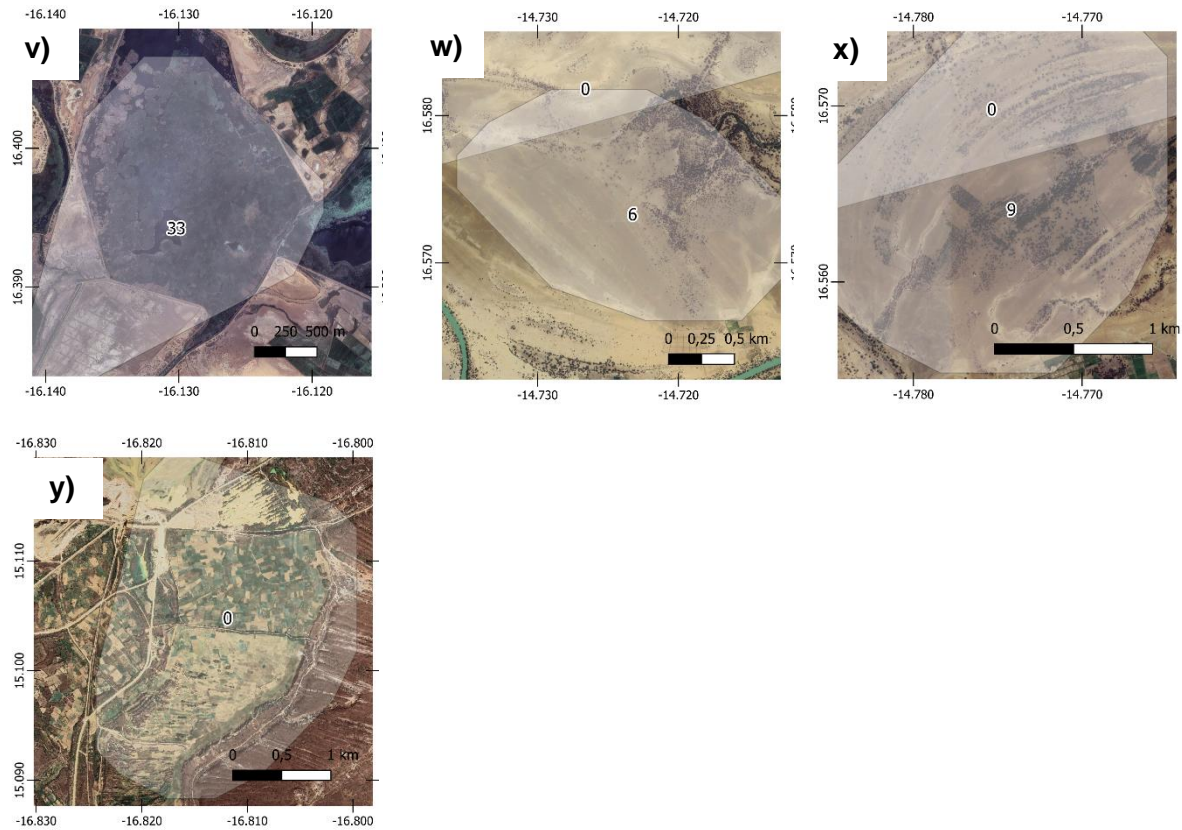


Figure 4.8: Close-up views of some of the identified change objects (whitish polygons). LSAs: subplots a to o (Niayes: a-f; SR: g-o). Infrastructures: p,s (Niayes). Natural (wooded land cover change): t (Niayes). Mines: q (Niayes). Urbanization: r (Niayes). Wetlands: u (Niayes), v (SR). Floodplains w-x (SR), Smallholder agriculture: y (Niayes).

**Figure 4.9** provides simple statistics on the number and area of the objects detected per land use type and study area. In the Niayes, about 1/3 of the detected objects correspond to wetlands located in the southernmost part of the study area (37), while another 1/3 is shared between LSAs (15) (some not in the field database) and infrastructure and urbanization (15). The remaining third is divided between mines (7), mixed (i.e. non-homogeneous) objects (3), natural areas (3), and small-scale (but dense) agriculture (1). With the exception of the mixed objects (most often a mixture with urbanization), the largest detected objects correspond to infrastructures (airport, road construction...), followed by LSAs (median of 276 and 165 ha respectively).



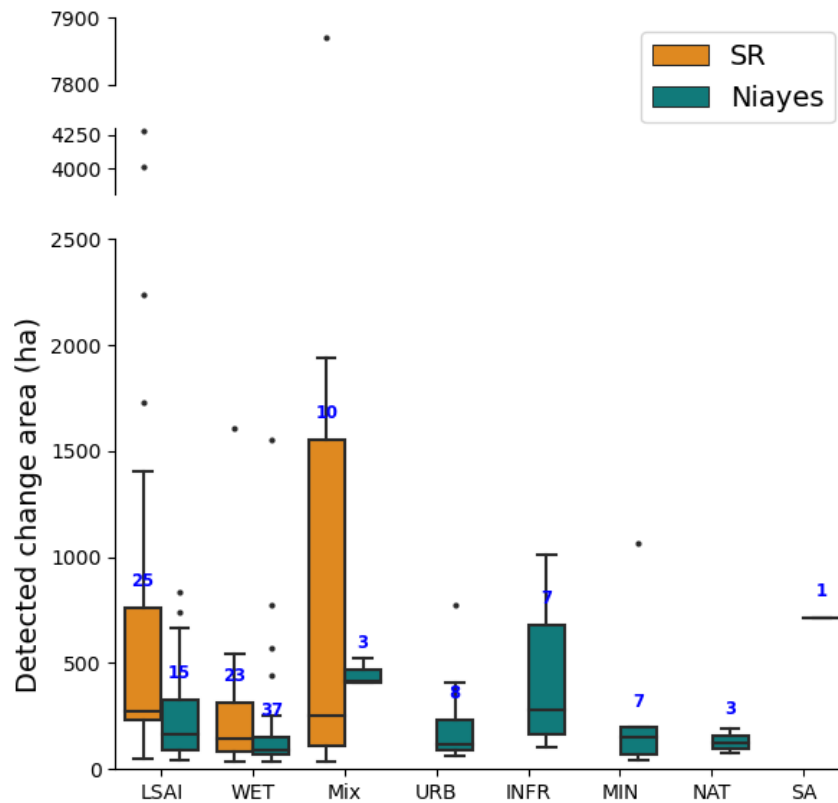


Figure 4.9: Box plots of the size of detected objects per land use type and study region versus their area. The number of observations in each class is shown in blue.

In the agriculturally intensive SR region, about half of the **58 detections** correspond to LSAIs (25) (some of which are not in the field database) (**subplots g-o of Figure 4.8**), followed by wetlands (this category also includes floodplains, see 2.3.3.1) (23) (**subplots v-x of Figure 4.8**). The remaining objects correspond to mixed objects (10). In this region, and with the exception of the mixed objects (often related to agricultural activities within floodplains), the largest objects are related to LSAIs (median = 274 ha), followed by wetlands/floodplains (median = 142 ha).

Now that the major hotspots of change have been extracted, the next question is to identify the major driver of change behind them. For this purpose, several object-based discriminative features were computed in an attempt to better distinguish LSAI from other land dynamics, in particular from wetlands/floodplains. These are presented in the next section.

#### 4.5.2.2 Spectro-temporal features

The discriminative power of various spectral, textural, and structural object-based features is evaluated in the following sections.

Pairwise scatter plots and distributions of each object-based calculated feature per LULC class are shown in **Figure 4.10**, separately for the Niayes (a) and the Senegal River region (b).

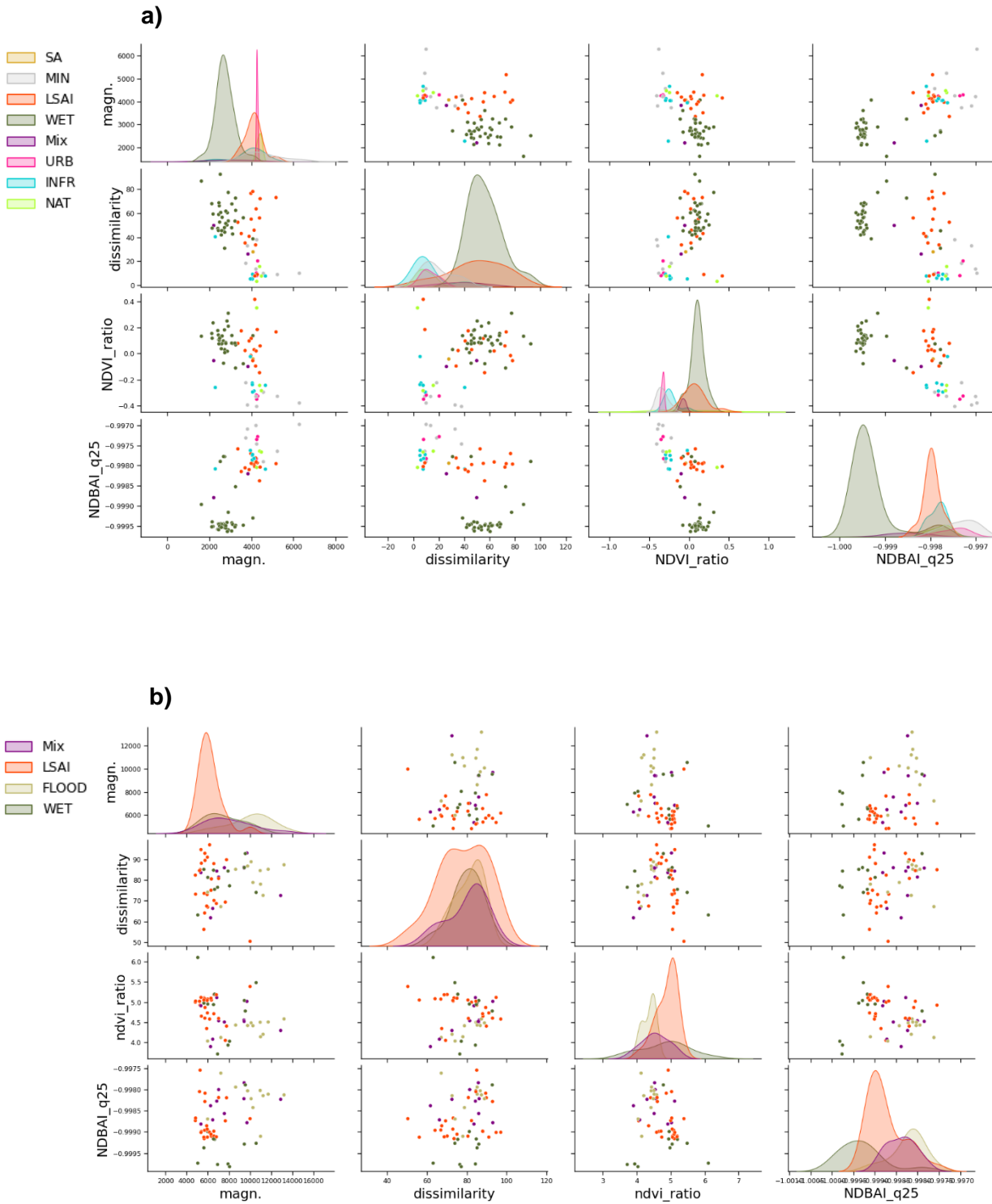


Figure 4.10: Pairwise scatterplots and univariate distributions (diagonal) of the spectro-temporal variables aggregated at the object-level, per LULC class for a) the Niayes, and b) the SR: the BFASTm-L2 magnitude of change, the time series shape dissimilarity, the ratio of the 3-year NDVI average before and after the detected change, and the NDBAI 25<sup>th</sup> percentile.

From **Figure 4.10**, several findings can be drawn. **The first is that the changes induced by LSAs are particularly similar to those induced by wetlands.** While LSAs can be distinguished from wetlands in the Niayes, this is hardly the case in the SR. In the Niayes, LSAs are well separated from wetlands by magnitude of change alone, with magnitudes above those induced by wetlands (~4000 vs. ~3000). In the SR, the opposite is true, with LSA-induced magnitudes (although higher than in the Niayes) below those induced by wetlands and especially floodplains (~6000 vs. ~8000 and 10000, respectively). **Second, the dissimilarity and NDVI ratio metrics allow good differentiation of biomass removal anthropogenic drivers** (mines, infrastructure development, urbanization) from wetlands and LSAs. Third, while the 25th percentile **NDBAI is particularly efficient at discriminating wetlands in the Niayes** (mostly mangrove type), it is less efficient in the SR where wetlands are mostly herbaceous (marshes). However, in this region, the NDBAI-based variable is useful in discriminating some of the induced changes in the floodplains.

In addition to these spectral variables, efforts were placed on the identification of textural and structural features that may be helpful in the discrimination of LSAs. These are presented in the next section.

#### 4.5.2.3 Textural and structural features

##### 4.5.2.3.1 Landsat NDMI composites

To extract textural and structural features, **the first step was to select the best image**, i.e. the image with the highest contrast between the detected object and its surroundings, which best represents the growing season. For this, the 2019-2021 NDMI images were first averaged on a monthly basis (thus obtaining 12 monthly averaged images). Then, a 3-month rolling average was applied. The image with the highest contrast was then selected. For illustration, the 3-year monthly NDMI average for February for one of the hotspot of change detected in the Niayes (**Figure 4.11.a**) is compared with the 3-month rolling average (**Figure 4.11.b**).

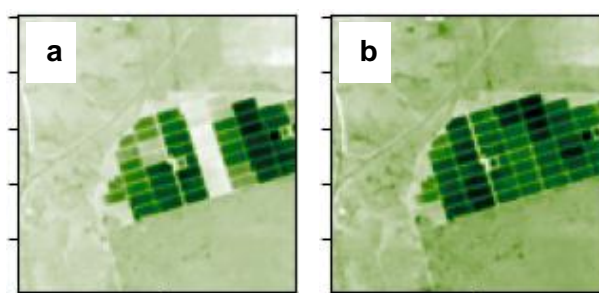
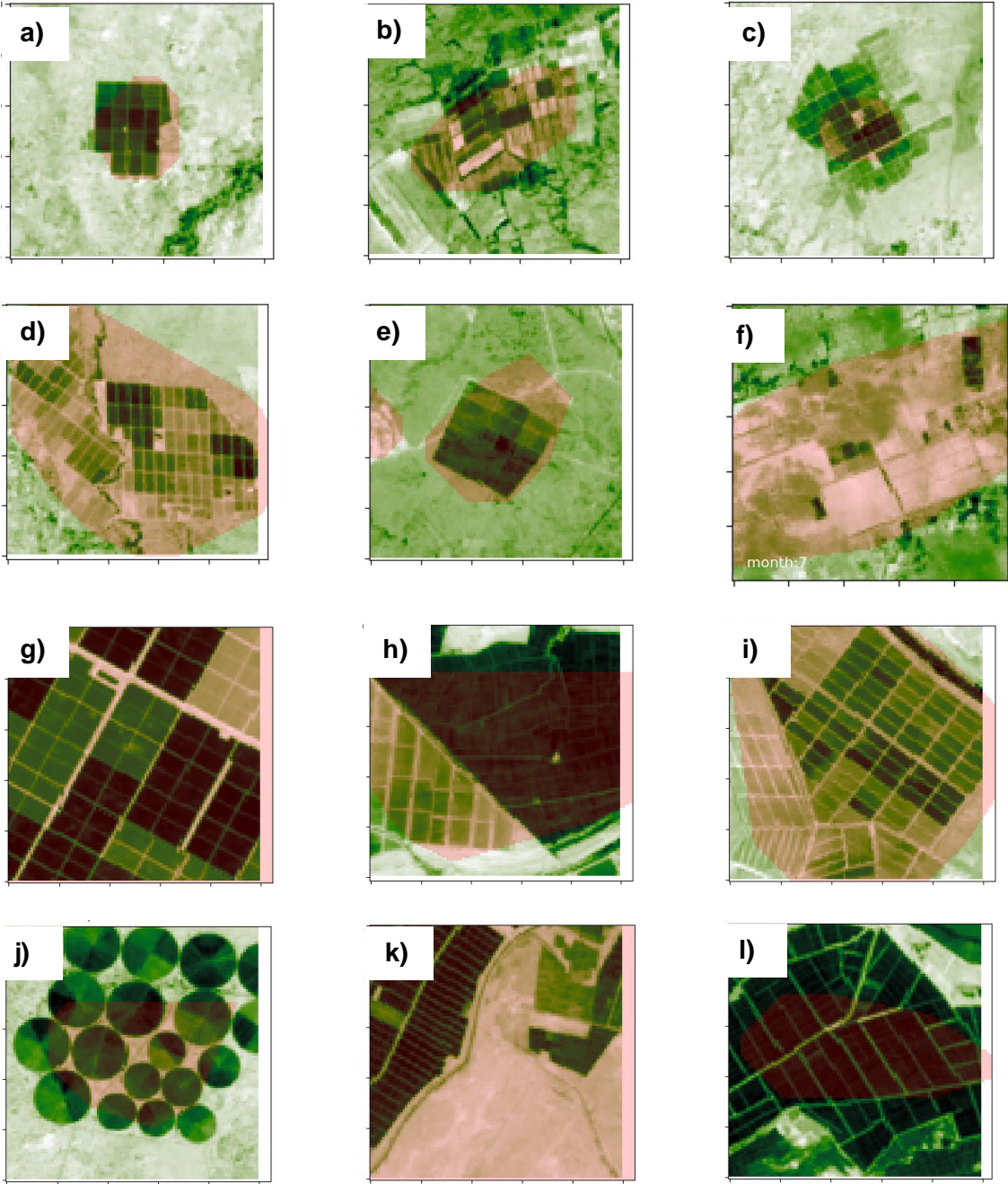


Figure 4.11: NDMI monthly means for February based on 2019-2021 LANDSAT data (normalized between 0-255). a) 3-year monthly mean; b) 3-year 3-month rolling mean.

As can be seen in Figure 4.11, the 3-month rolling average (b) gives a better picture of LSAI during the growing season than the 3-year monthly average (a). These composite NDMI-based images were therefore automatically selected and used as the basis for texture and structural calculations. **Figure 4.12** gives an overview of the obtained NDMI composite images for the same sample of various types of changes shown in **Figure 4.8**. As one can see, this compositing and selection method allows to obtain images where LSAs and agricultural plots are well contrasted (**subplots a-o of Figure 4.12**).





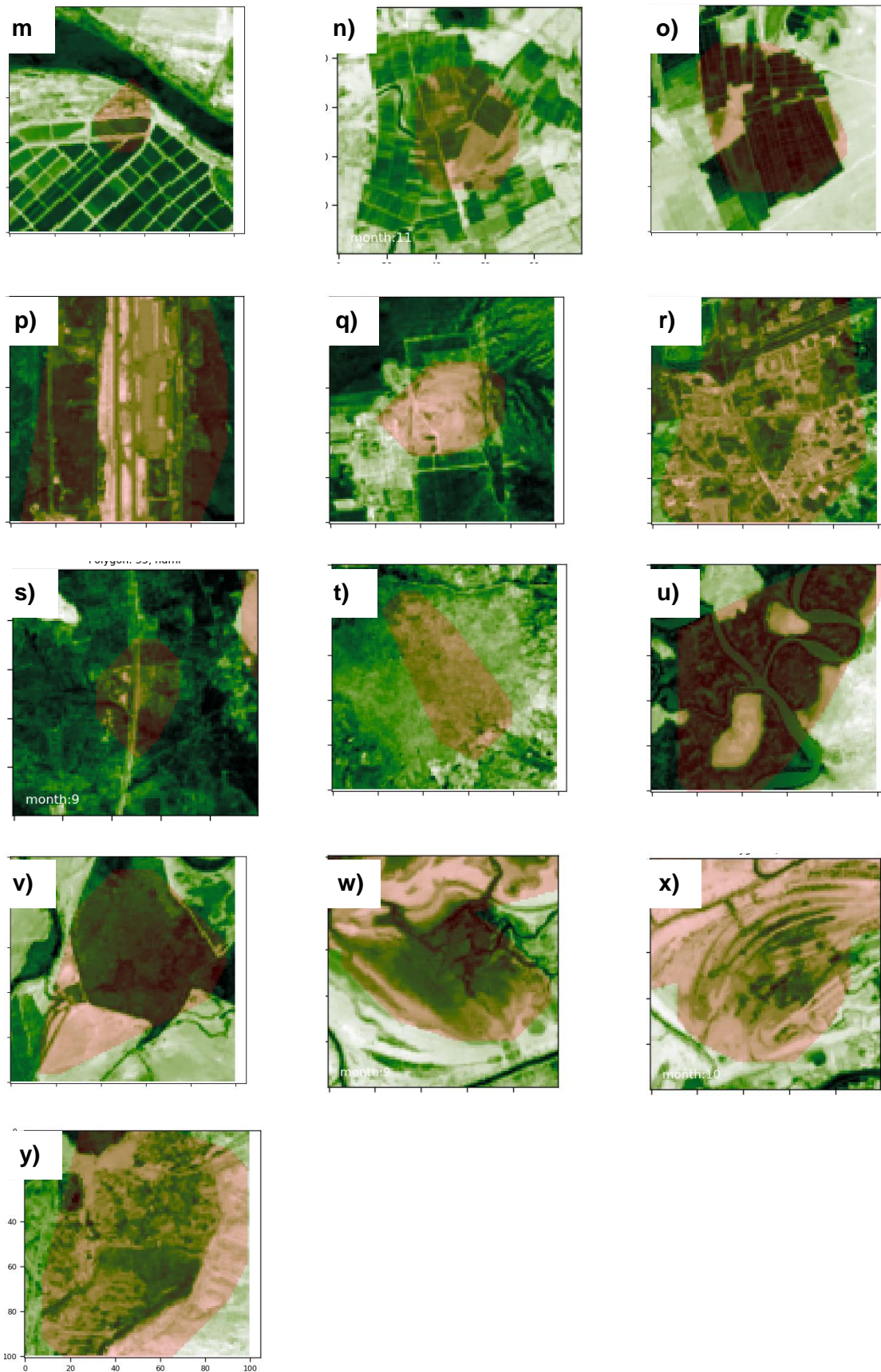


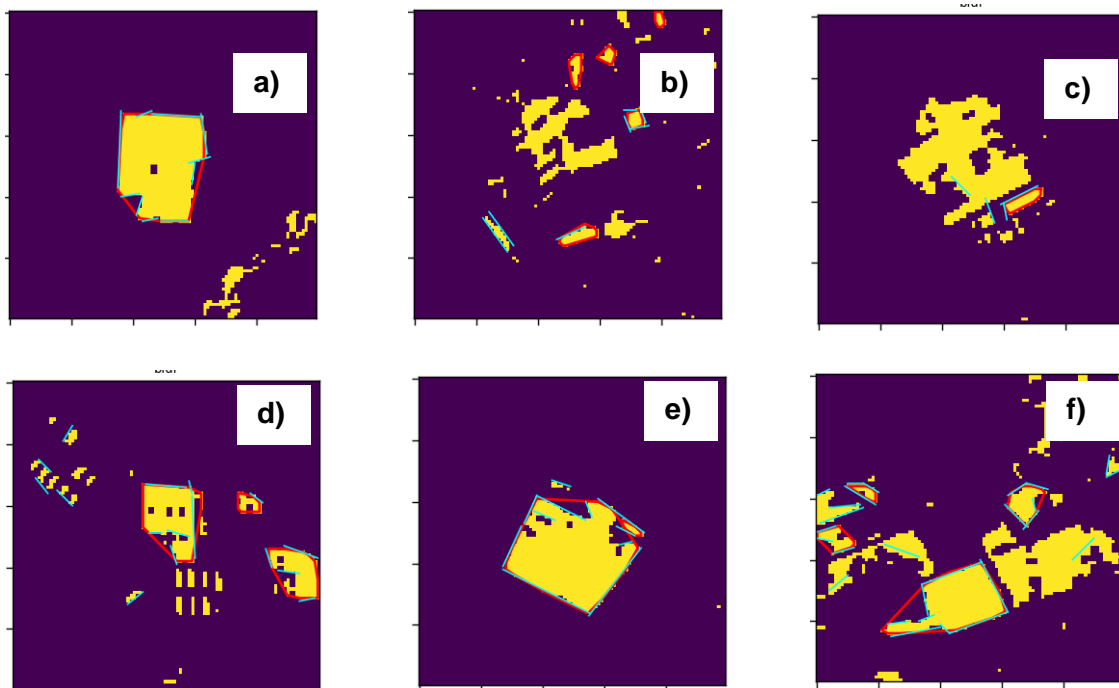
Figure 4.12: NDVI Landsat-based 3-year composites (100 x 100 pixels), centered at the hotspot of change, for different LULC changes in the Niayas and SR. a-o: LSAI, p: airport (INFR), q: mining (MIN), r: urbanization (URB), s: road construction (INFR), t: natural change (NAT), u-w: wetland (WET), x: floodplain (FLOOD) y: smallholding agriculture (SA). The light red masks indicate the MODIS-based hotspot of change.

4.5.2.3.2 Textural features

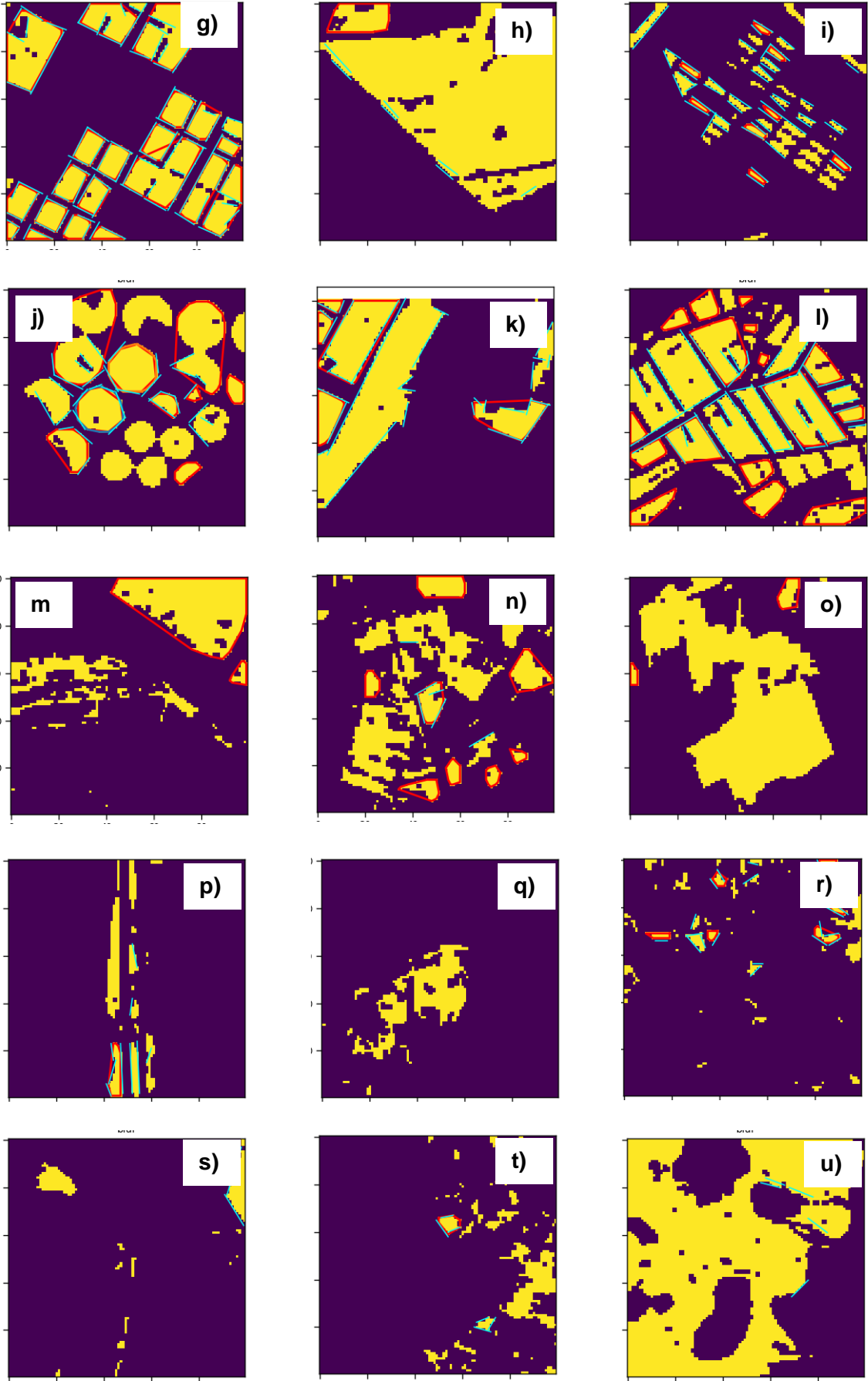
As mentioned above, the texture features were computed on the NDMI composites. A first observation from **Figure 4.12** is that the spatial **texture observed within the areas of change (in red) is very variable between the different cases**, as the plot size, plot shape, spatial arrangement of the plots (including the width between plots), and contrast to the immediate surroundings are very different for each case. As such, there is no clear common textural feature that emerges to distinguish these objects from others related to other drivers of change (subplots p-x in **Figure 4.12**). This observation is reflected in the distributions of the 14 Haralick's textures presented in **Appendix 4.A** (the Niayes) **and B** (the SR region) per type of LULC change driver. It can be seen that the separability between the different classes based on the textural features is low, especially between LSAI and WET. The Sum of Average (feature #6) is the feature that best separates the two classes in both regions, but there is still a significant overlap. Therefore, **these features are not included in the clustering step**.

4.5.2.3.3 Structural features

As for the textures, the structural features were extracted using the NDMI 3-month rolling averaged Landsat composites as a basis. Images were first pre-processed to optimize the image segmentation (see Section 3.2.2.2). The segmented images obtained (after image thresholding, see 4.4.3.4.2) for the samples presented in **Figure 4.8** and 12 are shown in **Figure 4.13**.







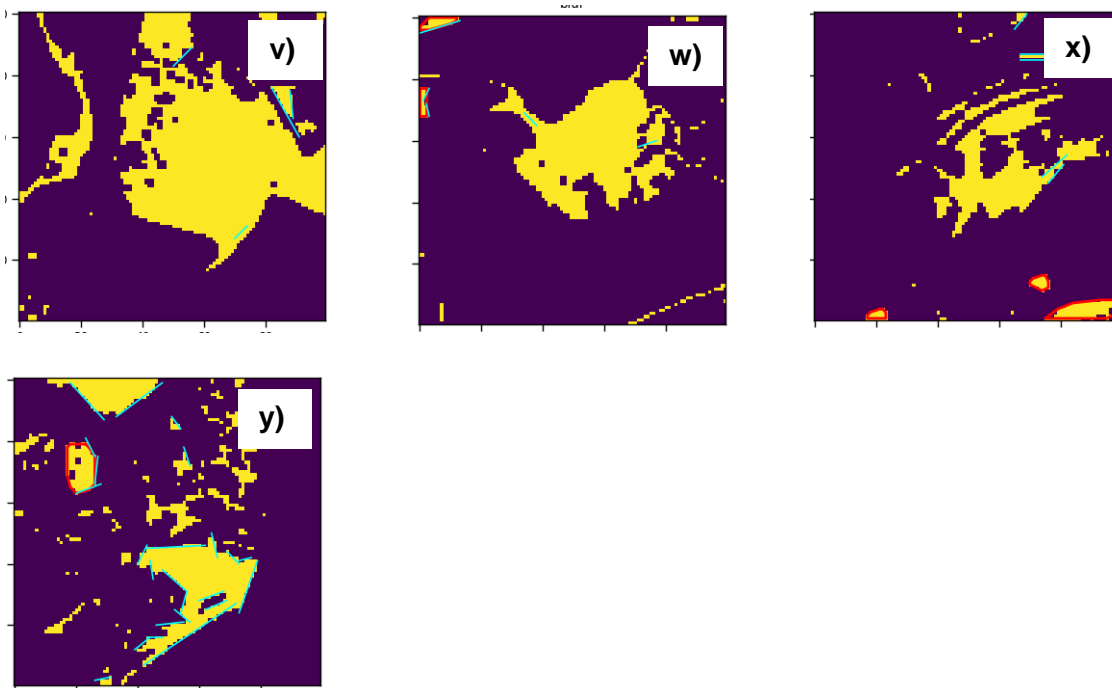


Figure 4.13: Segmentations (after image thresholding) obtained using the NDMI composites shown in Figure 12. Red polygons and cyan lines correspond to the detected and selected polygons/lines using the contour method and *lsd* algorithm respectively as described in section 2.3.3.2.

**Figure 4.13** shows that objects retrieved from LSAI-related imagery (subplots a to o) are overall less complex, with more geometric shapes, than those retrieved from imagery with other LULC types. This is truer for larger LSAIs in the SR (subplots g to o) than in the Niayes (subplots a to f). The more complex objects are obtained in locations where the LSAI parcel arrangement and size are not uniform (cases n and o), or where there is high within-field heterogeneity (crop type: case b). The contour-based method is overall efficient for detection of rectangular and circular LSAI plots. The evaluation of the *lsd*-based method used to detect lines is visually more difficult (note: only lines within the hotspot of change are retained). This is more easily done using the univariate distributions of the three calculated structural metrics shown in **Figure 4.14**, namely the cnt counter (i.e., the number of rectangular polygons/circles), the line counter (i.e., a number related to the number of parallel/perpendicular lines), and the geom index (an index that combines the cnt counter and the line counter, see 2.3.3.2). The distributions in **Figure 4.14**, show that in both regions, LSAIs have higher structural metric values than the other types of landscape objects. Overall, the combined metric, i.e., the geom index, is more efficient in differentiating LSAIs in the SR region than in the Niayes, especially from wetlands (see **Appendix 4.C**). As noted above, the contribution of the contour-based method is higher in the SR than in the Niayes, where the *lsd*-based method is more efficient.

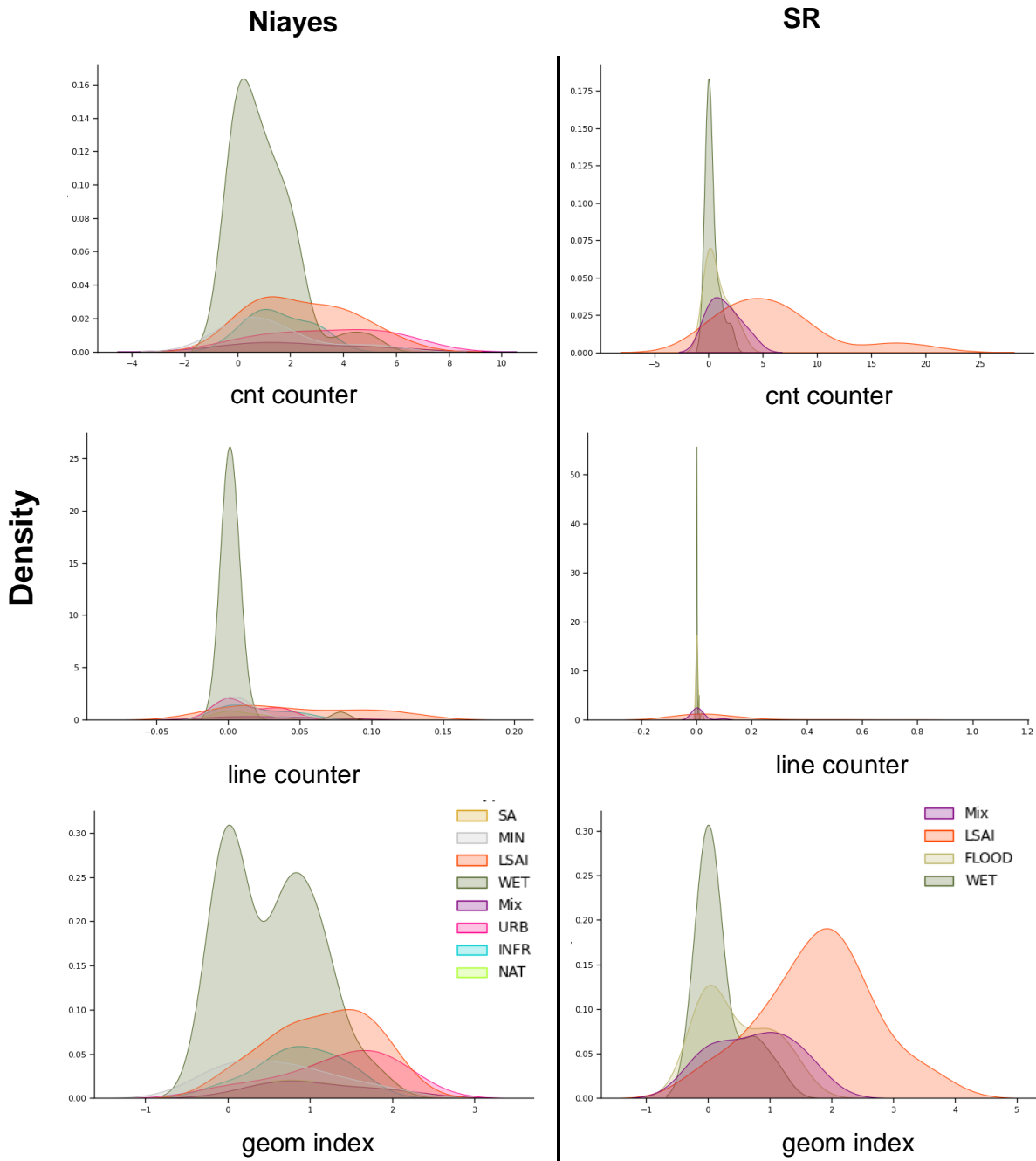


Figure 4.14: Univariate distributions of three structural features in the Niayes (left) and the SR region (right): 1: cnt counter, the number of square angles and circles detected using the contour method; 2: line counter, the number of parallel lines and squares angles detected using the lsd algorithm, 3: logarithmic variable combining both: geom index.

As observed, the structural features have a high discriminative power in some regions. However, due to their sensitivity to different parameters (i.e., preprocessing and detection methods), they are not included in the clustering step.

### 4.5.3 Clustering of change hotspots

Following the feature analysis performed in the previous section, a K-means clustering of the object based spectral features was performed for each region independently. For each region, the best partitioning (k value) was identified using the Silhouette metric as described in Section 3.3. Silhouette values for each clustering from k=2 to k=10 are shown in **Appendix 4.D**.

#### 4.5.3.1 The Niayes

**Figure 4.15** shows the partitioning of the 72 hotspots using k=5 in the Niayes region. This number is less than the total number of hotspots found (81), because of the unavailability of Landsat Tier1 images for some polygons. The 9 lost polygons belong mainly to the URB and WT classes. The overall silhouette width obtained was 0.5, indicating a reasonably good clustering overall.

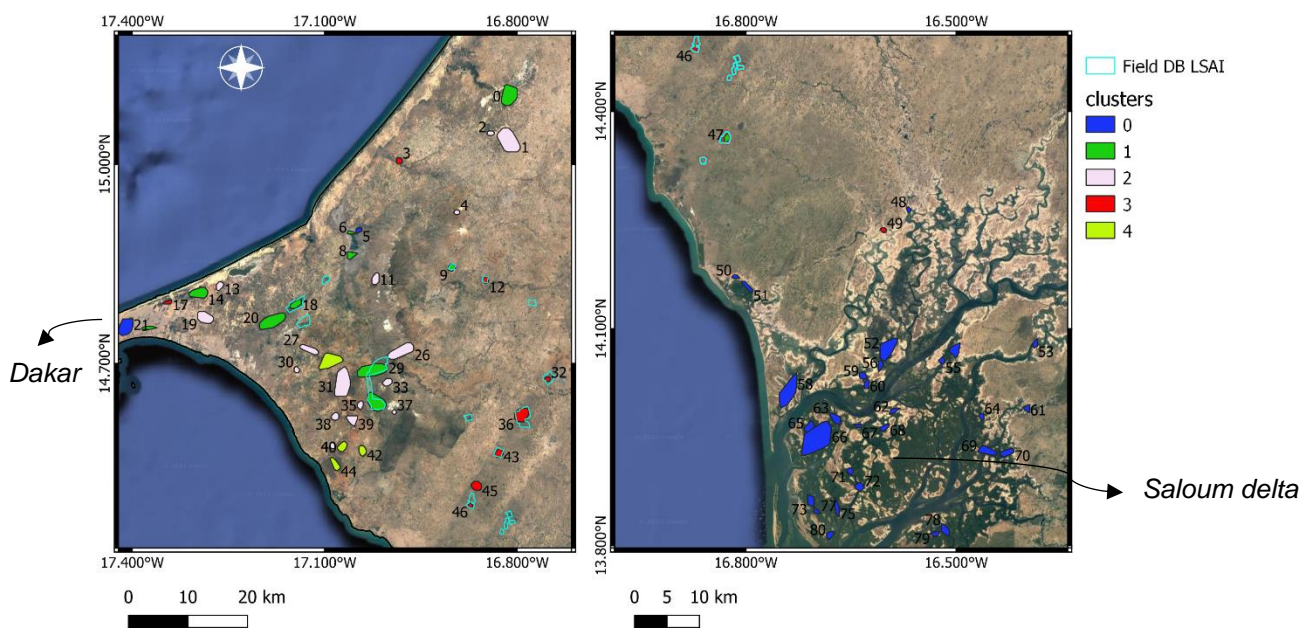


Figure 4.15: K-means clusters (applied to the extracted hotspots) obtained in the Niayes region with k=5.

The composition of each cluster is shown in **Figure 4.16**. The vegetation biomass removal LULC classes, i.e.: **URB, MIN and INFR have been combined into a new class ANTH (for anthropogenic, in pink)** for better clarity. In addition, a distinction was made for visually identified LSAIs that were not reported in the LSAI field database ("No-DB LSAIs"). These correspond to LSAIs that are active but were installed before 2003 (outside our change detection monitoring period) or have not been reported.

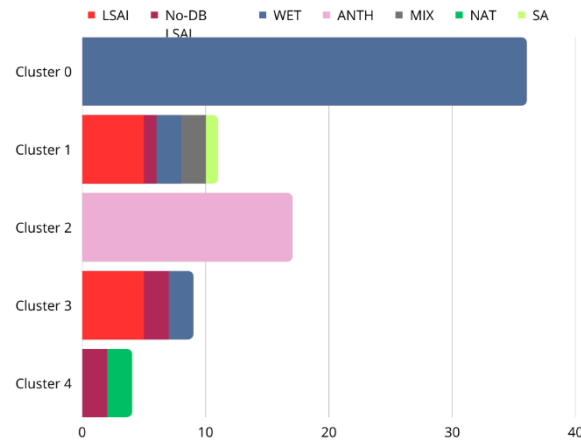


Figure 4.16: Barplots of the K-means cluster composition for the Niayes, with the x-axis the number of objects.

**Figure 4.16** indicates that the clustering of the spectro-temporal features in the Niayes allows a good differentiation of the wetlands (of mangrove type in the *Saloum delta*) and the vegetation removal anthropogenic drivers, represented by clusters 0 and 2, respectively. This is reflected by their high silhouette values (0.6 and 0.5 respectively, see **Appendix 4.D**). LSAIs (15 hotspots out of 72 are LSAIs) mainly belong to two clusters: cluster 1 (6 hotspots out of 11) and cluster 3 (7 out of 9). Both have a Silhouette average of 0.3, indicating a high within-class variability. WET objects within both clusters (in blue) are located near Dakar. These are different (less forested) from those found in the southern Saloum delta (cluster 0).

To understand what makes up each cluster, a boxplot of each clusters against the object derived features is shown in **Figure 4.17**.

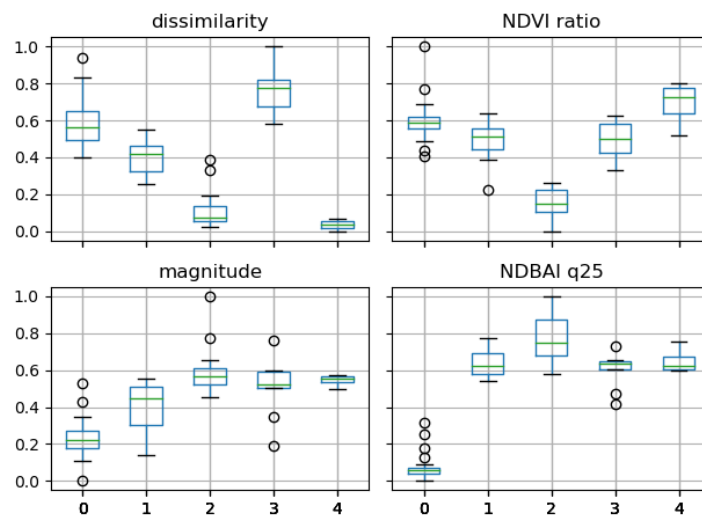


Figure 4.17: Boxplot of clusters (x-axis) in the Niayes against original spectral-temporal variables (one subplot per variable; values on y-axis).

Starting with the non-LSAI clusters (clusters 0 and 2), it can be seen that cluster 0 (wetlands) is mainly defined by very low NDBal values (characteristic of covered soils), but also by high seasonal variability (i.e. high dissimilarity and NDVI ratio). The associated BFASTm-L2 magnitudes are the lowest in the dataset. Cluster 2 (containing INFR, MIN and URB objects) is defined by the lowest NDVI values, indicating vegetation removal. This is supported by the highest NDBAI values, which are characteristic of bare soils. The changes observed within this cluster, which are among the highest-magnitude ones, are associated to very low dissimilarity values.

**LSAIs-related clusters (mostly clusters 1 and 3, but also cluster 4) share relatively high NDBal and NDVI ratio values.** Cluster 3 is different from the other two clusters in that it has extremely high dissimilarity values (the highest in the dataset) indicating a strong seasonal change. Cluster 4 has the highest NDVI ratios with the lowest dissimilarity values, indicating non-seasonal positive changes, most likely in amplitude or gradual. Finally, cluster 1 represents the one with the smallest magnitudes of change.

### 4.5.3.2 The Senegal River (SR) region

**Figure 4.18** shows the partitioning of the 57 change hotspots using  $k=7$  in the SR region. In this intensively agricultural region, the overall silhouette value obtained is 0.36 (Appendix 4.D), indicating an overall poor fit. The composition of each cluster is shown in **Figure 4.19**. As in the case of the Niayes study, a distinction was made between LSAI objects reported in the LSAI field database ("LSAI") and those visually identified as not ("No-DB LSAI").

In terms of cluster composition (Figure 4.19), and in contrast to the Niayes, there is only one very small pure class cluster, Cluster 3, with objects on the shore of Lake Guiers (which appears to be composed of wetlands). Regarding the other clusters, three are of particular interest as they cover most of the LSAIs: clusters 2 (12 objects out of 14), cluster 0 (7 objects out of 10) and cluster 6 (7 objects out of 11). As can be seen in **Figure 4.18**, cluster 2 is the cluster with the most overlap with the LSAI field database objects. Other LSAIs are detected that are not in the field database because they have a communal land tenure system despite their large size (these are part of a government project that has made large tracts of land available to small farmers, with communal irrigation systems). **Cluster 0** is specific in the sense that it includes the only two detected objects that represent pivot irrigation structures (objects 52 and 57), but also most of the objects in the floodplains that represent some agricultural activities (e.g., object 18 in **Figure 4.8.n**). Finally, **cluster 6**, the only of the three that has a good silhouette average (0.5 vs. 0.3), has predominantly objects within LSAIs, and some wetlands.



## Chapter 4: Results

Besides these LSAIs- related clusters, two clusters mostly contain only WET/FLOOD objects: **cluster 1** mostly constituted of FLOOD objects and **cluster 5**, mostly constituted of WET objects.

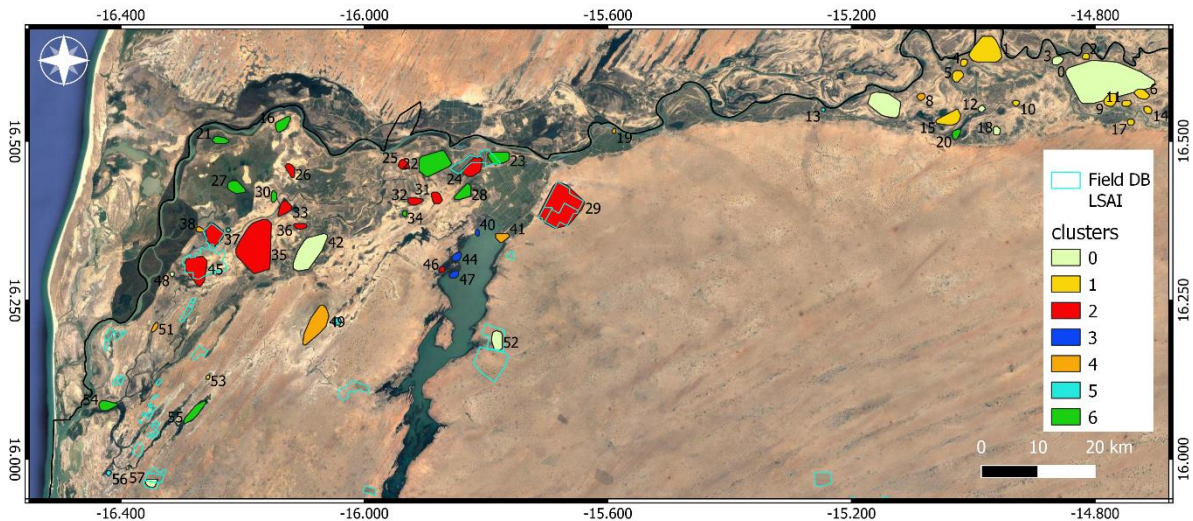


Figure 4.18: K-means clusters (applied to the extracted hotspots) obtained in the SR region with  $k=7$ .

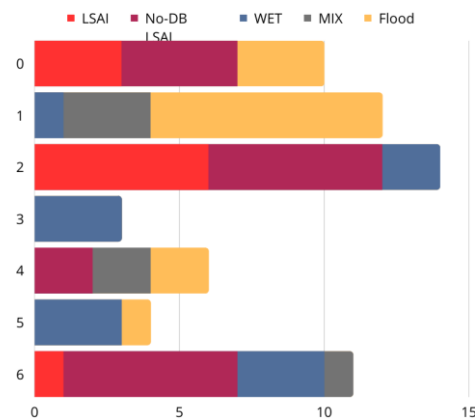


Figure 4.19: Barplots of the K-means clusters composition for the SR, with the x-axis representing the number of objects.

For better understanding, boxplots of each cluster against the object-derived features are shown in **Figure 4.20**. In this region, the clusters most associated with LSAIs, i.e. **clusters 2 and 6**, are mostly defined by very high NDVI ratios, very low magnitudes of change, and low NDBAI values. Compared to cluster 6 that is characterized by very high dissimilarity values, cluster 2 that contains most of the newest LSAIs is characterized by low dissimilarity values.

In the floodplains, the objects of change mostly belonging to **clusters 0 and 1**, are easily distinguished by their very high NDBAI and dissimilarity values. Compared to cluster 0, cluster 1 is mainly differentiated by its very high magnitudes of change (highest in the dataset).

Finally, remains clusters 3, 4 and 5. **Cluster 5**, mostly comprising wetlands (see above), has a strong seasonal variability, as indicated by its high magnitudes of change, high dissimilarity values and high NDVI ratios. **Clusters 3 and 4** are characterized by very low NDVI ratios, suggesting vegetation removal. However, when looking at the NDBAI values, the drivers of change seem to be different: while the extremely low values of cluster 3 suggest submergence, the very high NDBAI values of cluster 4 suggest land degradation.

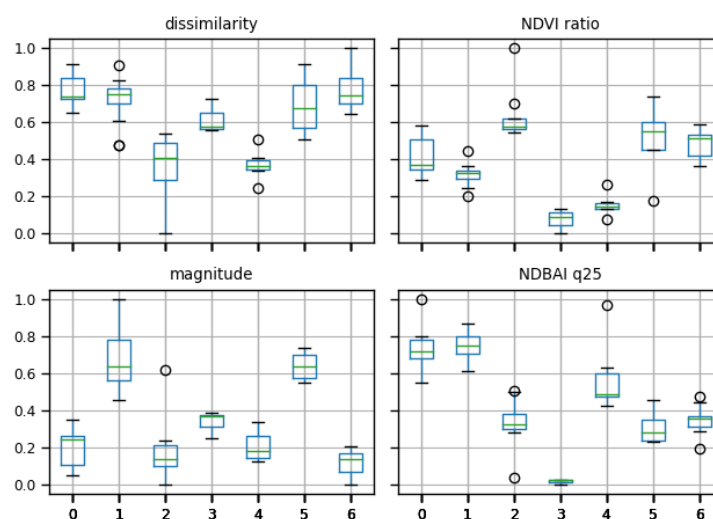


Figure 4.20: Boxplot of clusters (x-axis) in the SR against original spectro-temporal variables (one subplot per variable; normalized values on y-axis).

#### 4.5.4 Evaluation of the performance in detecting LSAs

As explained in Section 2.3.5, evaluating the detection performance of LSAs can be difficult, even with the availability of a field database that is not perfect and likely incomplete. In this study the evaluation consisted of two steps. In the first step, the number of LSAs detected, i.e. the number of change hotspots extracted using the BFASTm-L2 weighted magnitude map used alone, that overlap the LSAs reported in the field database (i.e., those implemented or extended within the monitoring period), was counted. This is given by the percentage of the first branched arrow in **Figure 4.21 a and b**: 53% for the Niayes and only 24% for the SR.

The second part of the evaluation is related to the discriminative power of the LSAs with respect to other land dynamics when performing a K-means clustering using the 4 spectro-temporal features. We want here to evaluate the precision of the "LSAI clusters": what is the probability of detecting a true LSAI when selecting these clusters? In this evaluation, we consider all the change hotspots visually identified as LSAI, even when not in the field database

("No-DB LSAI", in a burgundy color). The precision achieved for the LSAI-related clusters in the Niayes (clusters 1 and 3) was 65%, while in the SR it was 75% for the partition consisting of the combination of clusters 0, 2 and 6.

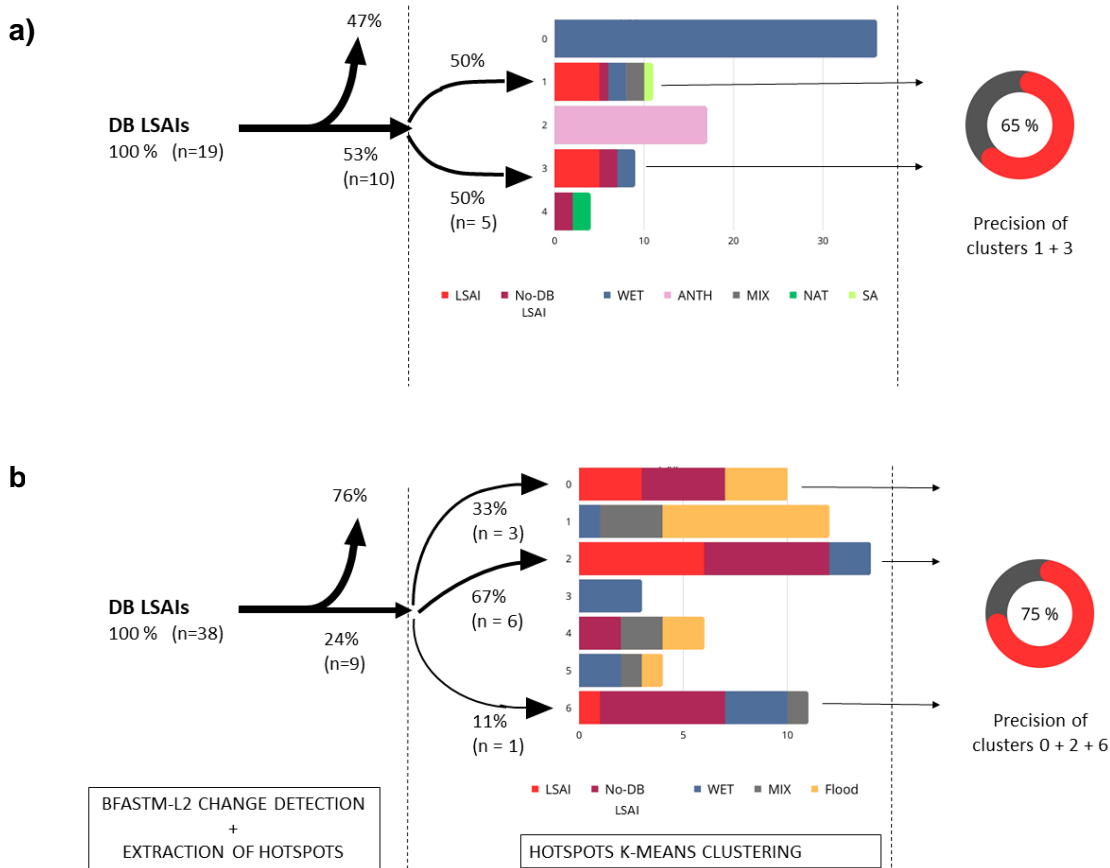


Figure 4.21: Tracking the destination of LSAIs reported in the field database through the steps of change detection (+ image segmentation) (first column) and hotspot of change clustering (2<sup>nd</sup> column) for a) the Niayes and b) the SR region. Precision of the combined LSAI-related cluster is given in column 3. Vertical arrows indicate the percent of LSAIs that were not detected. Multiple extracted hotspots of change may overlap the same LSAI, which explains why the sum of LSAI-related hotspots for the SR is greater than the number of DB LSAIs detected (10 vs. 9).

## 4.6 DISCUSSION

In this study, a three-step approach was implemented to detect and characterize LSAI-related LULC changes based on: i) BFASTm-L2 change detection on MODIS NDVI time series, ii) extraction of LULC change hotspots from a magnitude of change map weighted to highlight seasonal changes, and calculation of MODIS- and Landsat-based spectro-temporal, textural and structural features, and iii) K-means clustering of the change hotspots based on their spectro-temporal features, and identification and characterization of the most LSAI-related clusters. Since the BFASTm-L2 magnitude of change is somewhat sensitive to vegetation cover, the approach was independently applied, but using the same parameters, to an

extended region of the Niayes and another extended region of the Senegal River in the north, where a high presence of LSAs prevails. This in order to not encompass multiple distinct ecoregions.

### 4.6.1 Change detection and hotspot extraction

The first two steps of the approach allowed the extraction of hotspots of change that showed significant seasonal variation. The hotspots were found to often overlay specific landscape objects and were found to be spatially coherent and meaningful despite having somewhat imprecise contours, mainly due to the application of the convex hull transformation (**Figure 4.8**). Depending on the region, about one-fifth (Niayes) to one-half (SR) of the detected hotspots correspond to LSAI, demonstrating the ability of this approach to detect LSAI-related objects that have undergone multiple spatial expansions at different times, without the need to select specific satellite images for image segmentation, specific years for comparison, or the application of specific parameters. As a reminder, hotspots of change were extracted from a weighted magnitude of change map, which was intended to improve the highlighting of hotspots of change that included seasonal changes. While the use of this metric efficiently improved LSAI spatial contrast, particularly in the SR, where 60% of the objects had more than a 200% difference in magnitude between the object and its surrounding buffer, the results varied by region. In the SR, only 24% of the reported LSAs were effectively detected, compared to 53% in the Niayes (**Figure 4.9**). While these numbers may seem very low at a first look, it is important to keep in mind that these detections were made in complex and high-contrast regions where the target objects are relatively small, using the full MODIS NDVI SITS without the help of ancillary data or spatial filtering. In addition, the same parameters were kept for the two regions to keep the method as generic as possible.

The differences in results can be attributed to several factors. First, the Niayes are more contrasted in terms of land dynamics, with LSAs more dispersed across the landscape, while the SR is an intensively agricultural region, with LSAs close to other high weighted magnitude of change objects. In addition, a higher proportion of LSAs in the SR are relatively small, with 16% covering less than 50 ha, compared to 5% in the Niayes (**Appendix 4.E**). Second, the SR contains dynamic and unstable ecosystems, such as marsh-type wetlands and floodplains, that experience similar or greater rates of change, making it more difficult to extract between smaller objects. In contrast, the larger LSAs, with the exception of pivots, are better detected. In the Niayes, half of the LSAs are detected and extracted using the contour method, although challenges remain due to nearby areas of natural change of similar magnitude (forested areas) and cases where changes are not significant enough (some of these LSAs are dedicated to plantations, others have changes that are not persistent in time (abandoned?)) to induce

substantial magnitudes or dissimilarities. Of course, there are some cases of inaccurate detection with BFASTm-L2, but these have not yet been quantified. Conversely, it was interesting to observe that some, albeit few, LSAs were detected that were not originally included in the database. Therefore, conditions for effective LSAI detection with BFASTm-L2 were defined: the analyzed region should ideally consist of a single ecoregion (since the image preprocessing (i.e. percentile contrast stretching) may prevent the detection of local changes when integrating the extreme values that may occur in other ecoregions) , **the target change areas should be larger than 50 ha or far from unstable ecosystems, LSAs should involve fast-growing crops or significant field preparation, and induced changes should be persistent for at least 3 years.**

### 4.6.2 Insight and characterization of the drivers of change

Besides LSAs induced LULC changes, other land dynamics were detected, justifying the need to better characterize the LULC changes. In the extended Niayes region, about one-third of the detections corresponded to mangrove-type wetlands, another third to anthropogenic drivers related to vegetation removal (including infrastructure development, urbanization, and mining), and about one-fifth to LSAs. In the SR, about half of the changes are associated with marsh-type wetlands and floodplains, while the other half is associated with LSAs. The exploratory analysis of the extracted hotspots of change aimed to identify common spectro-temporal, textural, and structural features that may help distinguish the main drivers of change. These are discussed in the following sections.

#### 4.6.2.1 Spectro-temporal characterization

Spectro-temporal features analyzed included: the BFASTm-L2 magnitude of change, the dissimilarity, the 3-year post-change/pre-change NDVI mean ratio, and the NDBAI 25<sup>th</sup> percentile. These individual features, through K-means clustering **allow to differentiate particularly well mangrove-type wetlands and vegetation removal anthropogenic drivers of change:** i.e. infrastructure development (airport, roads...), urbanization, mines (**Figure 4.16**). They also allow for the detection of subtle differences that exist within the same class. What we initially considered to be a single class, e.g. WET, in fact comprises several subclasses, clearly distinguishing forested wetlands (mangrove type, low magnitudes of change) from marsh type wetlands (higher and LSAs similar magnitudes of change), floodplains (highest magnitudes of change), and even some cases of submergence.

The main postulate of this study, based on (Ngadi Scarpetta et al., 2024) (submitted for publication), is that large and persistent anthropogenic LULC changes induce important seasonal changes that can be assessed by the dissimilarity metric. This was most evident in the Niayes, where many LSAs induced very high dissimilarities (cluster 3, LOS/NOS seasonal

changes). The remaining LSAs (primarily in cluster 1), despite containing pixels with high dissimilarities, have moderate values once aggregated at the object-level. Several factors explain this: 1) the inclusion of pixels from nearby change areas with very low dissimilarity values (e.g. road construction) or from areas with no change as a result of the convex hull transformation, 2) pixels within the object with a shifted detected change date (for which the Euclidean value is the largest, but not the dissimilarity value), 3) pixels with moderate dissimilarity values despite the observation of a clear visual seasonal change. **This is the case for changes that mostly induce LOS (Length of Season) changes without significant NOS (Number of Seasons) changes (i.e. second crop cycle of small amplitude).** In the SR, it was noteworthy that the most extensive LSAs (belonging to cluster 2), implemented on nearly bare (very dry) soils, had some of the lowest object-based dissimilarity means (with the exception of LSAs with pivot structures, which had the highest dissimilarity values). This phenomenon seemed applicable under tropical conditions as well (Ngadi Scarpetta et al., 2024)(submitted for publication) and warrants further investigation. Conversely, changes occurring within pre-2003 existing LSAs induced very high dissimilarity values (see **Figure 4.20**, cluster 6), allowing for differentiation from newly implemented LSAs.

Compared to the dissimilarity, we observed that the magnitudes of change were highly variable as a function of region: while LSAs induced high magnitudes of change in the Niayes, they had relatively low values in the SR region compared to those induced by unstable ecosystems such as floodplains and marsh-type wetlands (see **Figure 4.20**, clusters 1 and 5). On the contrary, and as expected for arid regions, the NDVI ratios were relatively high for all the different cases.

In summary, the combination of these 4 spectral features has good discriminative power for differentiating LSAs. Since in both regions the LSAs were distributed in 2 to 3 clusters, we evaluate the precision of a combined and unique cluster for LSAs, which gives an indication of the probability of detecting a true LSA when using these 4 object-based spectro-temporal features. Despite the simplicity of the approach, the precision ranged from 65% to 75%, and in both regions the false positives were mainly represented by marsh-type wetlands. In order to achieve more accurate results, the next two sections discuss the appropriateness of adding textural and structural features to the analysis to better differentiate LSAs.

### 4.6.2.2 Textural characterization

In this study, global Haralick's texture features were extracted from the 3-year Landsat-based NDMI composites for each of the change objects. As can be seen in **Figure 4.12**, the LSAs are generally well contrasted, demonstrating the suitability of the compositing and image selection methods. However, and in contrast to studies where textural features have been



successfully applied (Bey et al., 2020; Kuemmerle et al., 2009), the textural features extracted here showed a very low discriminative power between the drivers of LULC change (as observed in the distributions per type of driver in **Appendices 4.A and 4.B**). This is mainly explained by the large variability observed in plot size, plot separation width and within-field intensity, making it very difficult to find a discriminating LSAI-related textural feature. In addition, while the use of an automatic but variable hotspot footprint can have some advantages, as it allows to reduce a large part of the spectral information not related to the LSAIs, it is not without its limitations, as the number of plots within the footprint can vary greatly (e.g. subplots g and m in **Figure 4.12**). This may further hinder the use of textural features as discriminative features. The most useful textural feature was found to be the *Sum of Average*. In this study, wetland discrimination was found to be particularly difficult (again, echoing the results obtained with the spectro-temporal features). The use of local texture features computed at different window sizes could potentially lead to better results. However, the automaticity of the method could not be guaranteed, since the optimal window size would be different for different cases, as evidenced by the large variability in plot sizes observed.

### 4.6.2.3 Structural characterization

The last group of features to be assessed is the structural one. The main objective here was to find some metrics that could correlate with the geometric shapes very often observed in intensive agricultural landscapes as a result of mechanization (Tang et al., 2021; Vogels et al., 2019; Yan and Roy, 2014). Two methods were tested: 1- based on the extraction of contours, 2- based on the extraction of lines. In both cases, the number of square angles and circles (contour method) or parallel lines (line method) was calculated. In both cases, image preprocessing constituted a very important step, and the results were very sensitive to the different parameters used. Therefore, while we manually adjusted the parameters to values that worked best for both regions, we decided not to include these metrics in the unsupervised classification at this exploratory stage. Here we found that in both regions the LSAIs show indeed an overall high number of geometrical shapes, compared to other detected landscape objects (except, and as expected, for the URB class in the Niayes). The contour method performed particularly well in the SR region, allowing to discriminate LSAIs from wetlands (see **Figure 4.14, Appendix 4.C**). This was favored by the higher contrast objects and the larger size of the parcels. In the Niayes, the lower contrast and more heterogeneous parcels found (see **Figure 4.14.b**), in addition to the smaller and more "compact" fields observed, which do not allow the extraction of individual field parcels (as in **Figure 4.14.g**), lower the structural metric values. In addition, in this region, more geometric shapes were found in the wetlands due to what appears to be the many sandbanks found isolated in the mangrove wetlands.

Adding these features to the unsupervised classification would therefore lower the precision of the clustering to LSAs in the Niayes, while greatly improving that in the SR region.

### 4.6.3 Limitations, recommendations and perspectives

The proposed approach to automatically detect LULC changes potentially related to LSAs has shown to be promising. The work presented here is mainly exploratory, and as such has certainly avenues worth exploring.

One key area for consideration is the scalability and transferability of the approach to different regions. While designed for large-scale applicability, our findings emphasize the importance of performing the spatial analysis in an area that does not encompass different ecoregions due to variations in change magnitudes and change characteristics. Extending the testing to different regions, particularly tropical ones, represents a logical progression, although challenges in acquiring sufficient high-quality temporal data may arise due to cloud cover.

Other avenues to explore are related to the metrics of change used. The main hypothesis tested in this study is that LULC changes associated with LSAI should mainly affect the seasonality of the time series, quantified by the dissimilarity metric integrated in the weighted change map and in the clustering. While this was indeed observed in the Niayes, the semi-arid region of the Senegal River, the most representative cluster of LSAs shows very low dissimilarity values compared to other land dynamics, especially those related to wetlands and floodplains. This is interesting because the induced change in the time series pattern can be significant. This has also been observed in subtropical environments, with significant changes (but in the opposite direction) likely related to field preparation (Ngadi Scarpetta et al., 2024). While these changes are very large, with significant changes in amplitude, they are often accompanied by changes in length of season (LOS), to which the dissimilarity metric is less sensitive. Therefore, the exploration of new dissimilarity metrics that are more sensitive to changes in LOS is desirable. This could also potentially improve change detection at the pixel level. Continuing with the metrics of change, the exploration of new metrics may also be of interest, particularly the inclusion of temporal metrics that allow characterization of the rate, distribution, and frequency with which some changes occur at the object level. This could potentially better characterize LSAs from other land use systems, such as smallholder agriculture.

When considering spatial-related features, our study suggests that incorporating structural features derived from high spatial resolution satellite imagery improves spatial analysis, as observed in the semi-arid Senegal River region. Simple features, such as square angles and circles, effectively distinguished LSAs from wetlands. Although the proposed approach is rudimentary, its flexibility (no search for perfect shapes) and simplicity are advantages.

Optimizing automatically the image preprocessing (although no hard thresholds were applied) and evaluating when the integration of structural features is beneficial (as there is a large variability in the spatial footprints exhibited by LSAIs) represent areas for refinement.

Finally, the question of applying the fully unsupervised approach to new regions remains. Although time consuming, identification of clusters associated with LSAIs may involve visual inspection. This should be feasible as the overall approach was thought to provide the lowest but still relevant number of hotspots of change and clusters. This was done to avoid complicating the spatial analysis. In addition, alternative clustering algorithms could be tested to better handle the overlapping multivariate distributions observed in this study. For example, hierarchical clustering algorithms, which allow a better evaluation of the similarity of clusters, would be interesting to quickly identify the most related clusters.

The proposed methodology for automated detection of land use and land cover (LULC) changes potentially associated with LSAI shows promise. This study is primarily exploratory and suggests avenues for further investigation.

### 4.7 CONCLUSIONS

In this exploratory study, an unsupervised LULC change detection approach is proposed, focusing on the detection of a specific land use system (LSAI for Large Scale Agricultural Investment) in two contrasting regions of Senegal. The method, which aims to be fast and generic, is based on the detection of pattern changes within long and dense MODIS NDVI time series using the BFASTm-L2 algorithm, followed by the extraction of change hotspots from a change magnitude map weighted to highlight seasonal changes. The extracted hotspots are further analyzed and characterized using spectro-temporal (mostly MODIS time series change metrics) and Landsat-based textural (Haralick's features) and structural (based on contour analysis and lsd algorithms) features. In each study region, LSAI-related K-means clusters were identified and characterized. The precision (to LSAIs) of the cluster resulting from merging all clusters containing LSAIs, , ranged from 65% to 75%. Although still at an exploratory stage, the integration of structural features to discriminate LSAIs shows promise, and suggests many avenues for further investigation and method improvement.

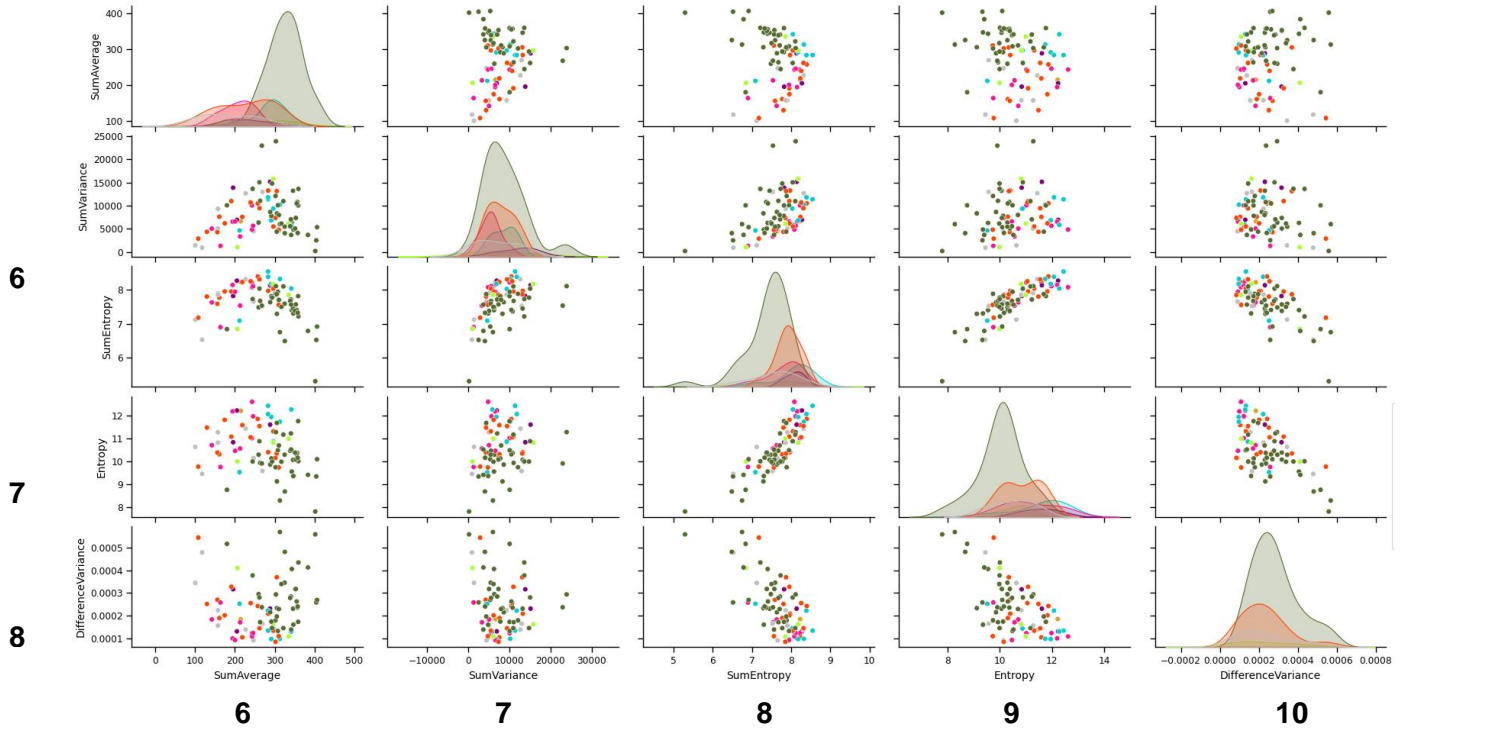
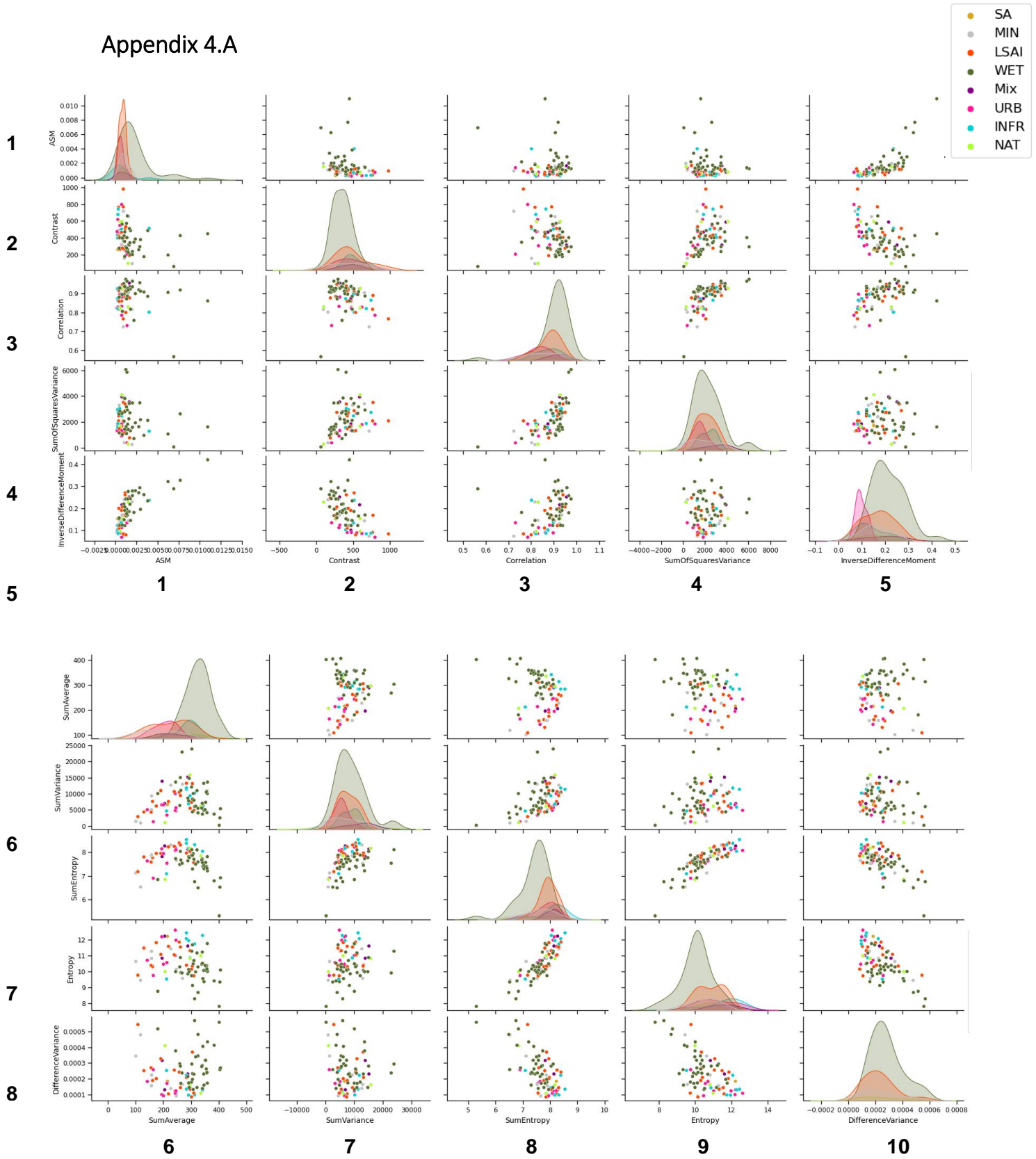
In conclusion, this approach provides valuable insights into LULC changes associated with LSAI and offers potential for further methodological improvements. This study was conducted with the specific purpose of assisting concerned citizens and stakeholders, or LSLA monitoring initiatives such as the Land Matrix ([landmatrix.org](http://landmatrix.org)), in identifying significant LULC changes that could potentially be attributed to LSAI. This approach serves the dual purpose of

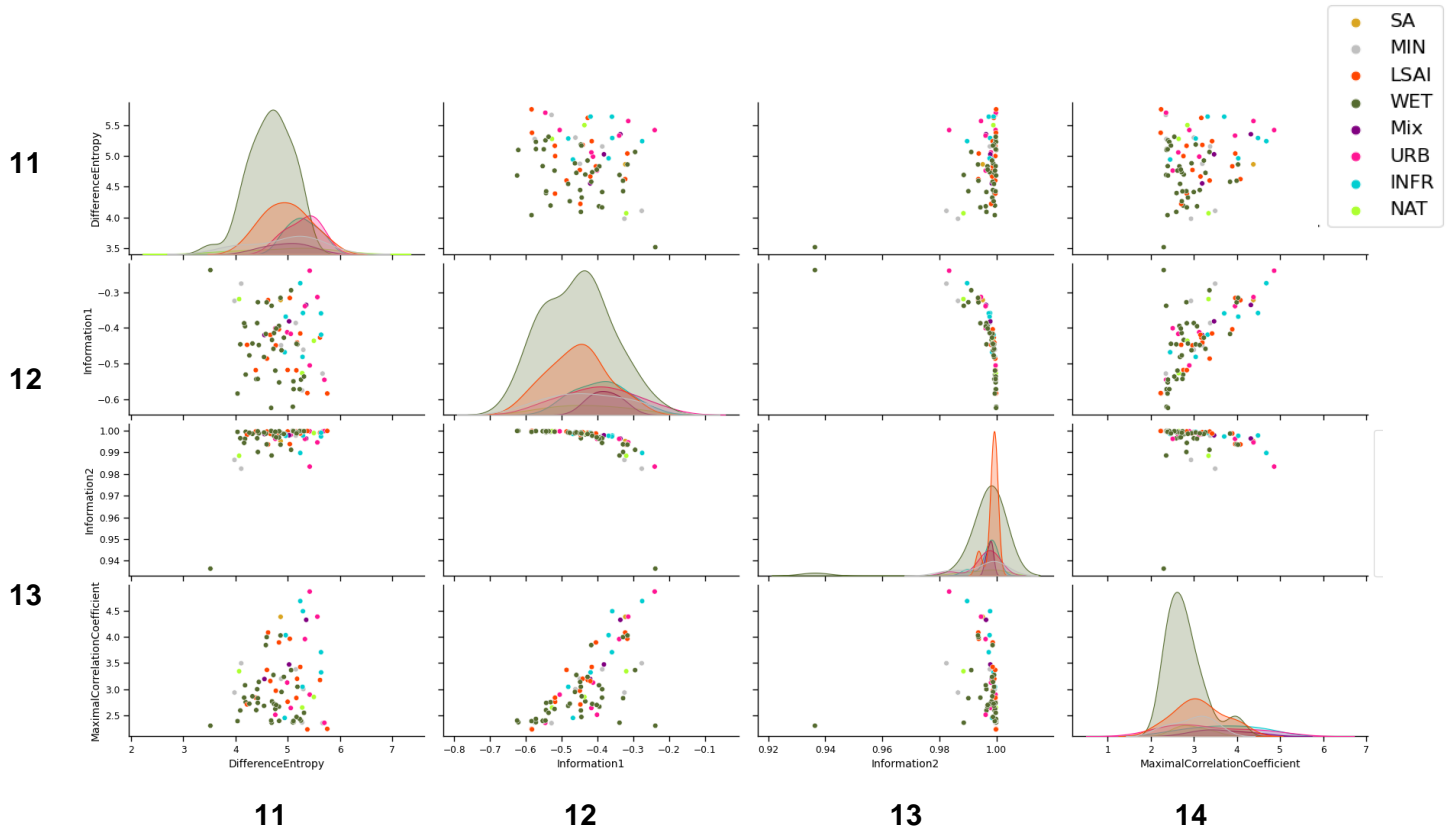
## Chapter 4: Conclusions

identifying regions of particular concern and facilitating the coordination of future field campaigns, improving transparency, and paving the way for impact analysis.

## 4.8 APPENDICES

### Appendix 4.A

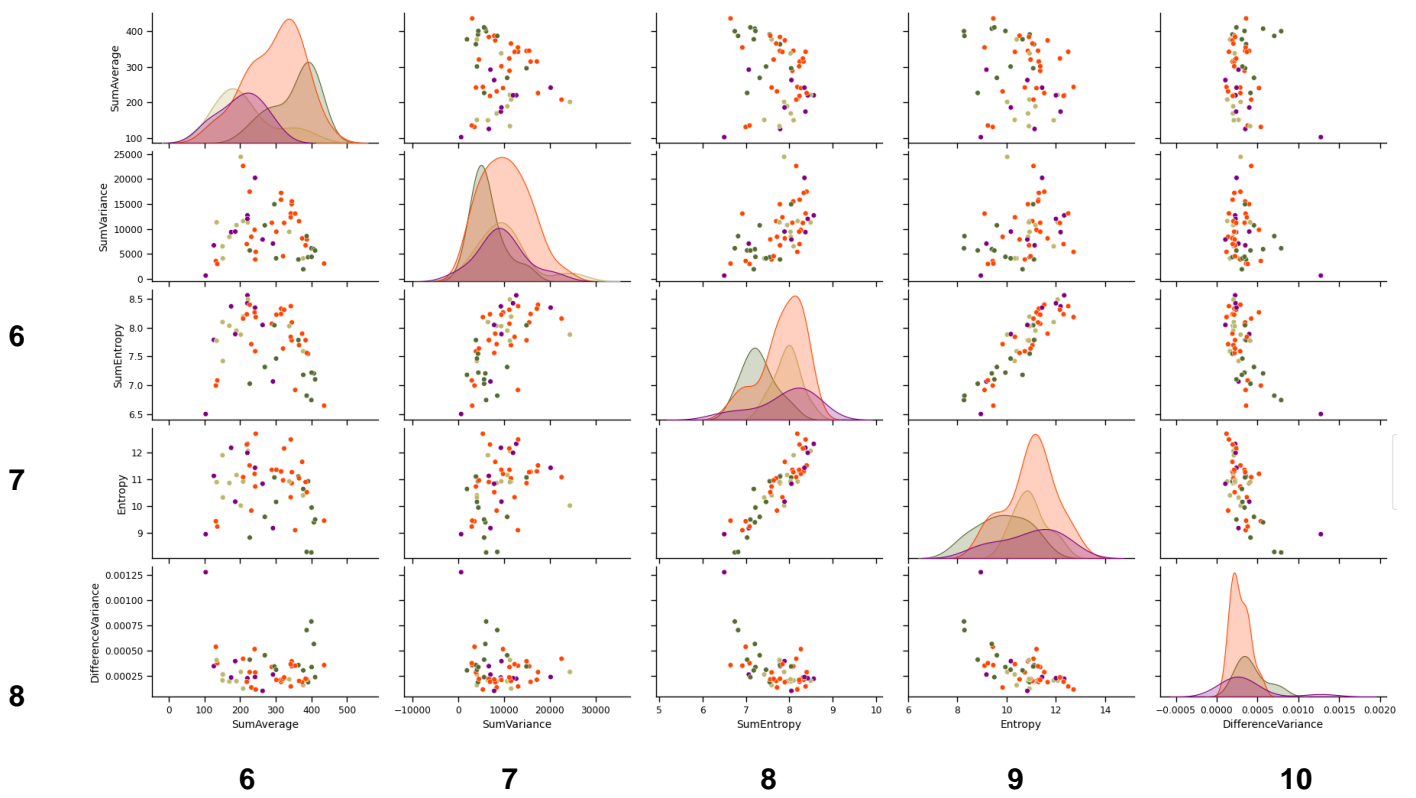
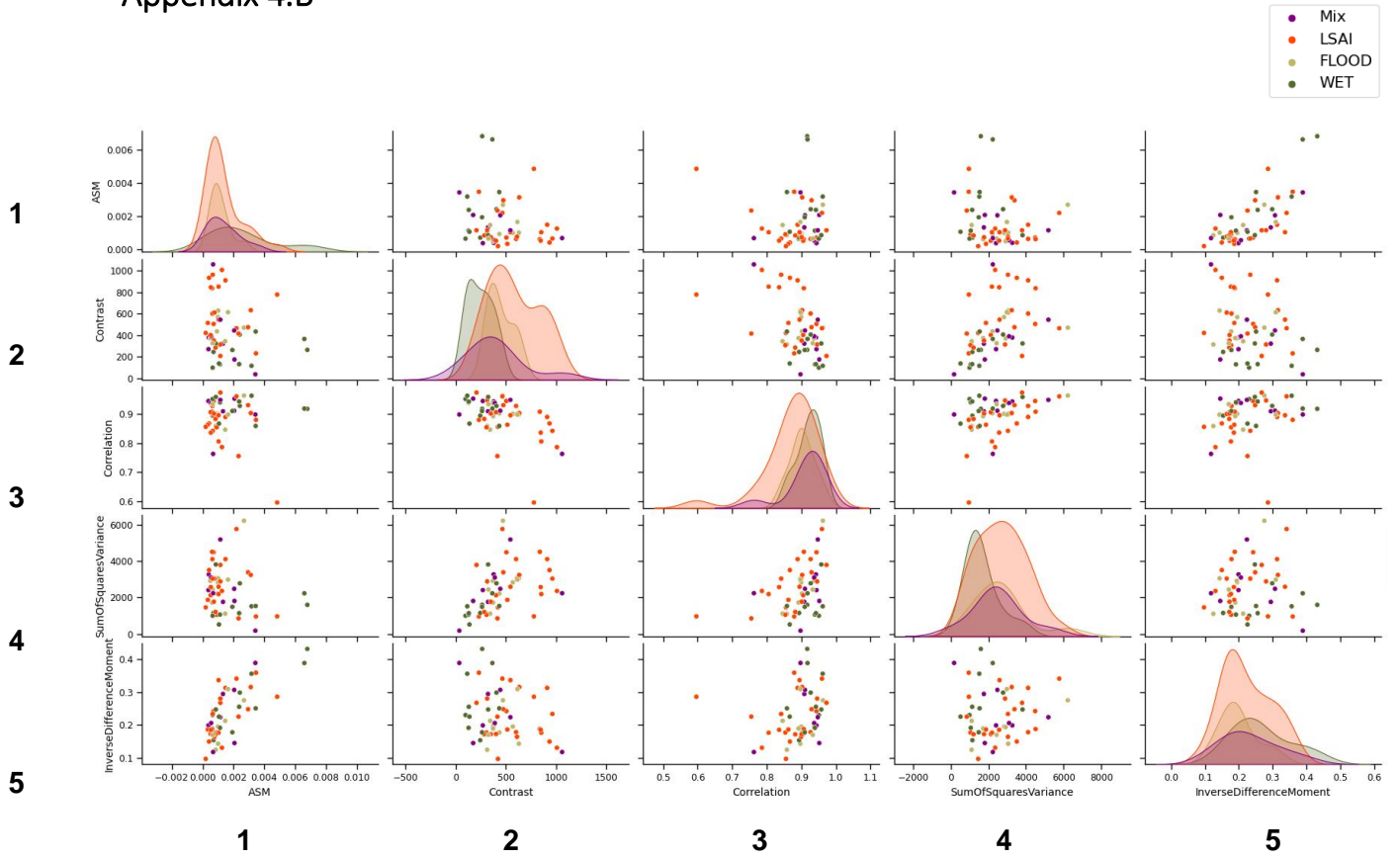


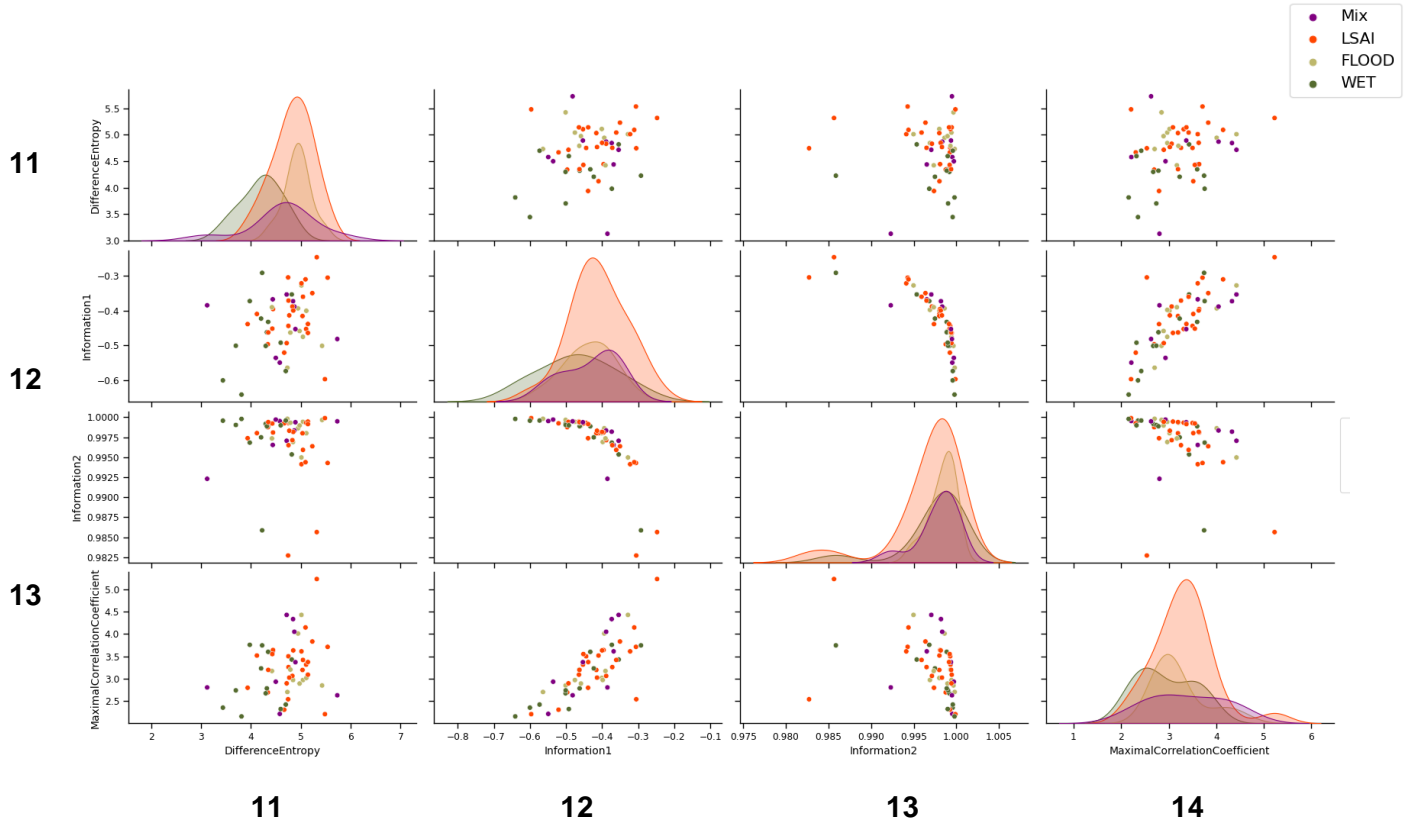


Pairwise scatterplots and univariate distributions of the Haralick's textural features for the Niayes: 1: ASM, 2: contrast, 3: correlation, 4: sum of square variance, 5: Inverse Difference Moment, 6 : Sum Average, 7: Sum Variance, 8: Sum Entropy, 9: Entropy, 10: Difference Variance, 11: Difference Entropy, 12: Information1, 13: Information 2, 14: Maximal Correlation Coefficient.



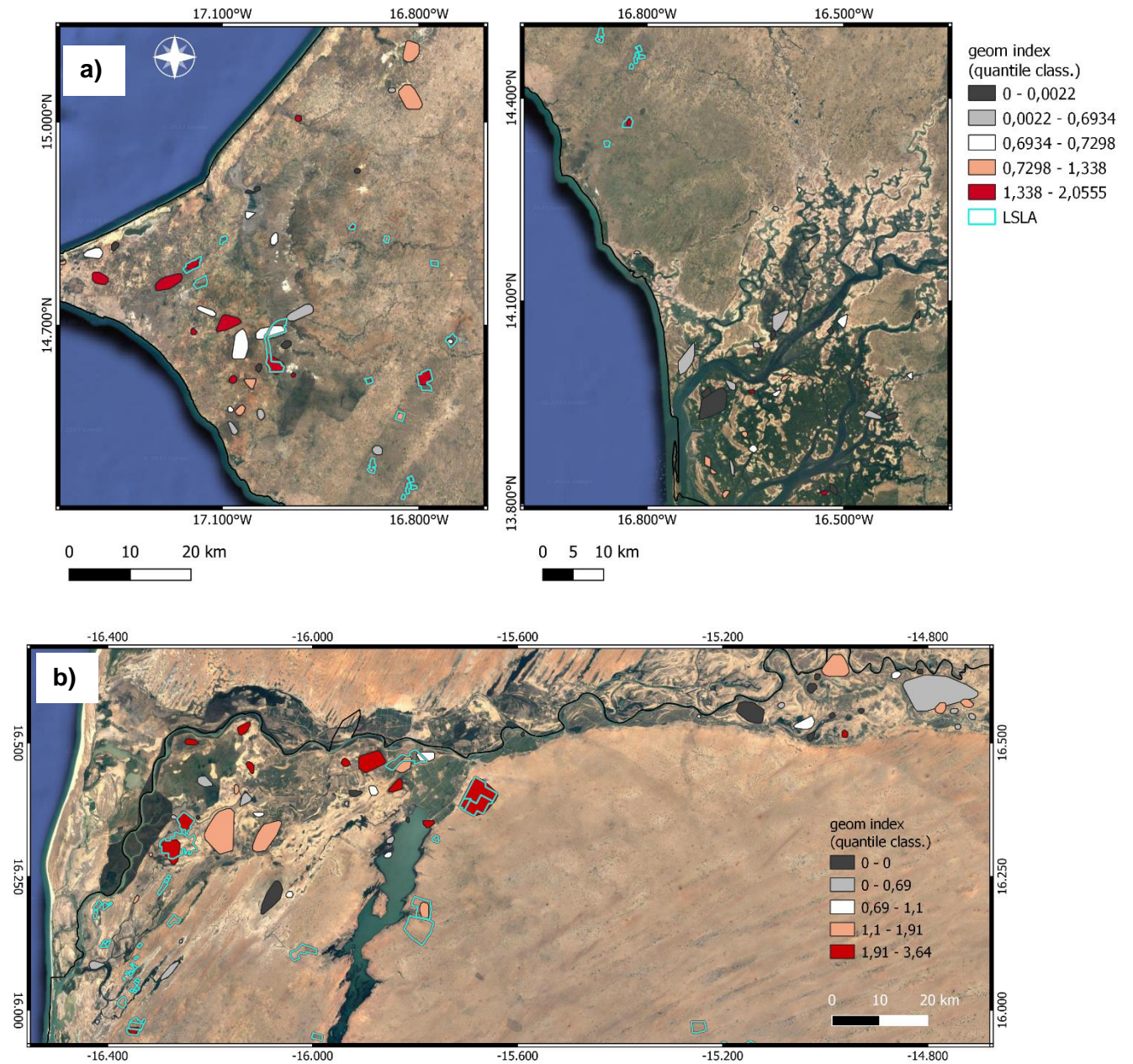
Appendix 4.B





Pairwise scatterplots and univariate distributions of the Haralick's textural features for the SR region: 1: ASM, 2: contrast, 3: correlation, 4: sum of square variance, 5: Inverse Difference Moment, 6 : Sum Average, 7: Sum Variance, 8: Sum Entropy, 9: Entropy, 10: Difference Variance, 11: Difference Entropy, 12: Information1, 13: Information 2, 14: Maximal Correlation Coefficient.

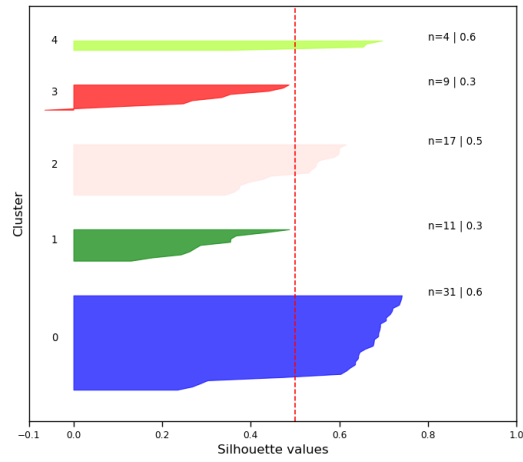
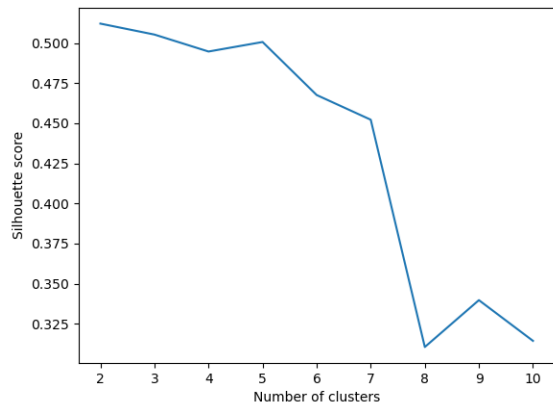
Appendix 4.C



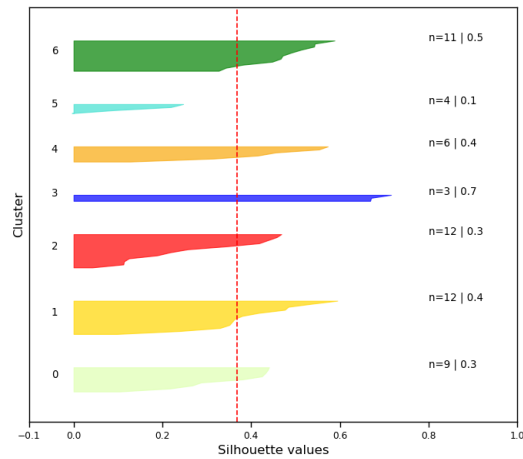
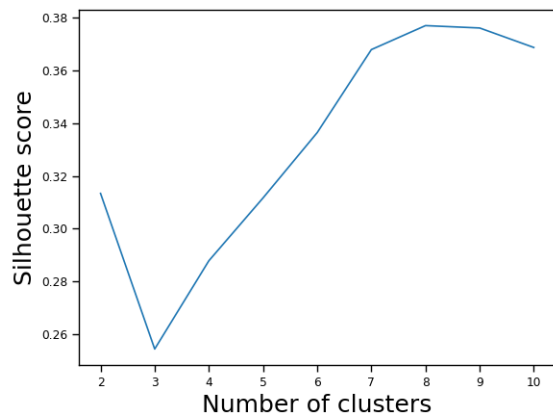
Classification of the hotspots of change by their geom index value for a) the Niayes and b) the SR. In both maps 5 classes were defined using a quantile classification. LSAIs reported in the field database are shown in cyan.

Appendix 4.D

a)

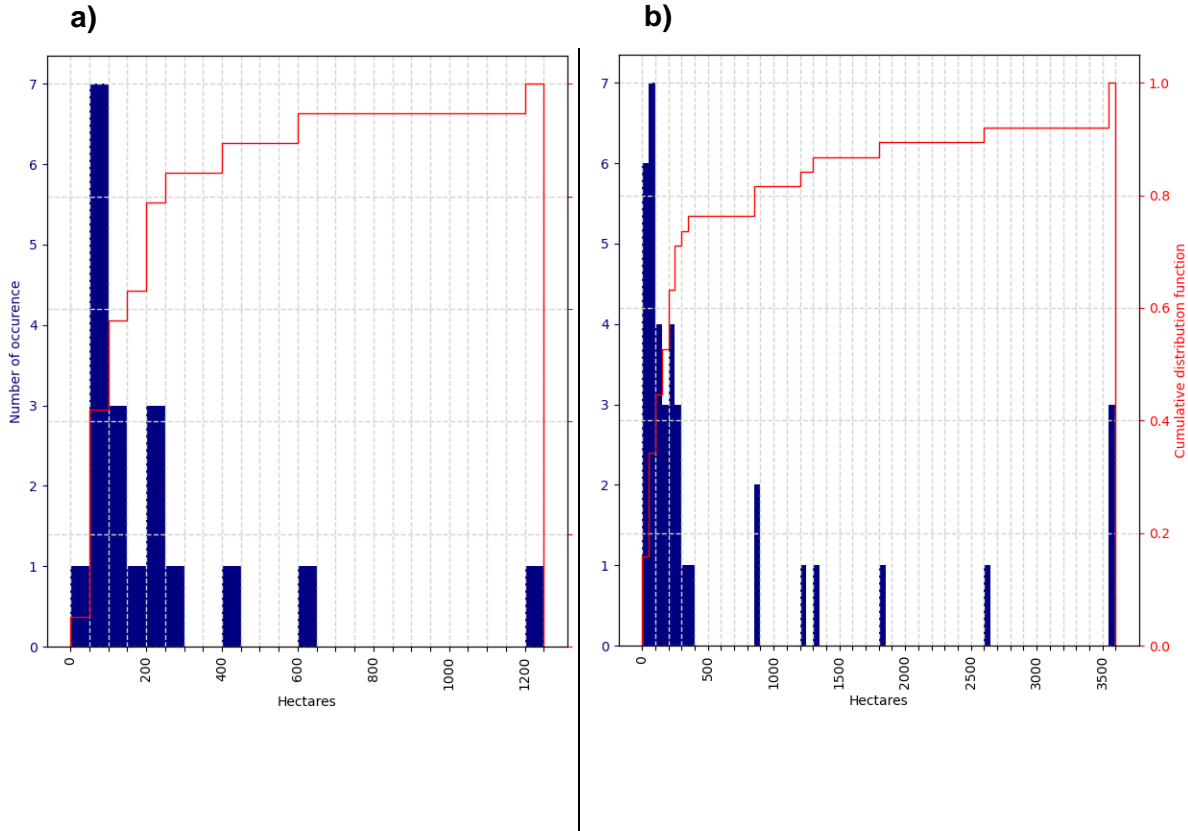


b)



*Silhouette plots: a) the Niayes, b) the SR region. Each cluster is plotted as a horizontal bar plot, with one bar for each object. Bar values near 1 indicate a good fit, while values near zero and below indicate a very poor fit with a high probability of misclassification. The overall average fit is indicated by the dashed red bar. The cluster size and the Silhouette average of each cluster are shown on the right.*

Appendix 4.E



Hotspot size distribution sfor: a) the Niayas, and b) the SR region.

## 5 DISCUSSION

---

This work has been driven primarily by the need for automated monitoring of specific land uses on a large scale for more effective resource management and better informed decision making. The question addressed in this research specifically concerns the detection of LSAI (Large-Scale Agricultural Investments) in tropical areas, as these systems are of strategic importance in terms of territorial development and socio-economic and environmental impacts, but remain poorly understood. The lack of reliable and up-to-date data on these systems at the national scale, due to the difficulty of collection, makes Earth observation data particularly well suited to address this issue. Since LSAs are complex land use systems with many different forms (in terms of crops, environment, practices...), the focus was on the development of an automatic, robust (i.e. generic) and interpretable method to detect LSAs at the regional scale. Therefore, the main question we addressed in this work was: **Can we automatically detect emerging large-scale agricultural investments (LSAIs) at a large scale without the need for a training dataset (i.e., in an unsupervised manner)?**

To address this overarching question, we have developed a conceptual model that links remote sensing features (spectro-temporal and spatial) to the characteristics of the land systems under study. Our strategy revolves around three sub-objectives, each of which is described in detail in the form of scientific papers in Chapters 2 to 4 of this document.

Chapter 2 (focusing on the first sub-objective) focused on the development of the BFASTm-L2 change detection algorithm. **This algorithm is designed to automatically detect significant pattern changes, primarily seasonal changes**, in the MODIS 250-m NDVI time series that may be associated with potential anthropogenic changes. We assumed that newly implemented LSAs are causing significant and persistent changes in vegetation composition and dynamics. Key requirements for the algorithm included unsupervised operation (i.e. independent of external data), efficiency for large-scale application (i.e. fast), sensitivity to seasonal changes, and robustness with minimal tunable parameters. In addition, the method had to avoid relying on rigid thresholds for either the magnitude of change or its direction (i.e., positive/negative change) to account for variation across ecoregions (e.g., arid vs. tropical humid conditions).

Chapter 3 (centered on the second sub-objective) aims to **understand the land use and land cover (LULC) changes likely to be induced by climatic, natural, anthropogenic (non-agricultural) and agricultural factors (including LSAs) at the national scale**. This step is critical for distinguishing previously detected LULC changes associated with LSAs. The proposed approach is grounded in the assumption that different drivers of LULC change exhibit



**distinct spectro-temporal signatures in terms of time series types of change.** Our strategy involved identifying three spectro-temporal features having contrasting sensitivities to different time series change types. Combining these features in an RGB composite map allowed highlighting different LULC changes with different signatures of change. By considering the area and shape of the observed changes, they could be linked to different drivers of LULC change.

The third and final sub-objective, outlined in Chapter 4, capitalized on the fact that LSAs induce large spatial changes, with large BFASTm-L2 change magnitudes that are combined to a shape dissimilarity metric to take into account seasonal changes, and particular landscape organization. Using the change map derived from BFASTm-L2, **the aim was to identify and analyze hotspots of change, integrating approaches and insights from the previous two sub-objectives, to pinpoint areas potentially associated with Large Scale Agricultural Investment (LSAI). Spectro-temporal evidence from 250m MODIS time series and structural/textural features from 30m Landsat imagery specifically computed from these change hotspots were used.** Although discriminative in some regions, structural features were excluded from the unsupervised classification analysis to maintain methodological robustness and genericity, as not all LSAs have distinctive geometric shapes (e.g., see the example of large rubber plantations in Laos in Appendix B.1, which were tested but not included in the research as a ground truth database was not available). The detection approach was independently tested on two contrasting sub-national regions in Senegal (arid and semi-tropical) with different types of LSAs.

These three sub-objectives are in line with the framework proposed by Zhu et al. (2022) for characterizing land change, which includes five facets of change, including *Where* (i.e., the location of change), *When* (i.e., the date of change), *What* (i.e., what is changing), *How* (i.e., the metrics of change), and *Why* (i.e., the driver of change). This discussion section summarizes the key findings through the lens of these facets of change, beginning with the *Where and When* facets and then delving into the subsequent *How* and *Why* facets that follow.

## 5.1 KEY FINDINGS

### 5.1.1 Where and When?

Chapter 2 enabled contributions to the *Where* and *When* facets of change with the development of BFASTm-L2, a fast and unsupervised change detection algorithm focused on identifying the breakpoint associated with the most significant spectro-temporal (i.e., mostly seasonal) change within long and dense MODIS 250-m NDVI time series. The underlying assumption was that newly implemented LSAs induce significant and persistent changes in

land surface vegetation composition that manifest as pattern (i.e. shape) changes within VI time series. In the process, we made three critical observations about existing algorithms:

- First, statistical change detection algorithms that emphasize speed often do not perform seasonal decomposition of time series, leaving ambiguity about the nature of the change being detected - whether it is abrupt, gradual, or seasonal, and if seasonal, how it affects seasonal components such as amplitude, length of season (LOS), or number of seasons (NOS). This makes it difficult to understand the land dynamics behind the detected change.
- Second, many algorithms tend to evaluate the magnitude of detected breakpoints based on deviations from the mean, which is appropriate for detecting abrupt or gradual changes but less effective for quantifying seasonal changes.
- Third, selection of the most significant breakpoint among multiple breakpoints, which is not always the largest abrupt change, is often desired but most of the time not provided.

The BFASTm-L2 change detection algorithm was developed in response to these challenges and compared with widely used algorithms (BFAST Lite, BFAST monitor, EWMACD) selected from the literature specifically for their speed and, where applicable, sensitivity to seasonal changes. Derived from BFAST monitor, BFASTm-L2 uses the Euclidean distance as the magnitude of change calculated between the two 3-year samples of the time series located on each side of the breakpoint detected by BFAST monitor. This distance has been found to be sensitive to pattern changes and in particular to NOS changes. It therefore allows the largest and most persistent (over three years) changes to be selected from multiple detected breakpoints.

**Key findings from these comparisons included BFASTm-L2's ability to spatially highlight known LSAs by mapping the magnitude of change (Euclidean distance), confirming that LSAI-induced LULC changes primarily affect the seasonal component (but not only, as explained in Box 5.1.1).** In addition, the observed mapping differences between the tested change detection algorithms were for some algorithms (BFAST Lite) due to differences in change magnitude rather than inaccurate temporal detection, confirming that the change magnitude calculated as the deviation from the mean may be inappropriate for quantifying seasonal change.

Having successfully addressed **Research Question 1.1 - What method allows detection and selection of the breakpoint that causes the greatest *pattern* change in long and dense NDVI time series** - the focus shifted to understanding the likelihood of detecting LULC changes driven by LULC drivers other than LSAI. This inquiry required a deeper understanding of the relationships between the major drivers of LULC change, such as climate variability,

natural factors, anthropogenic influences (including LSAs), and the resulting time series types of change. Chapter 3 dealt with this aspect, and key results are presented in the following section.

### **Box 5.1.1 – BFASTm-L2: Implications of detecting a unique breakpoint for LSAI detection**

BFASTm-L2 locates the breakpoint associated with the most significant pattern change (evaluated over a limited 3-year period). While this allows for the identification of persistent seasonal changes, it favors the detection of significant abrupt changes when present, a likely situation in real long time series. What are the implications of these findings? Although Chapter 2 showed that LSAs in the Niayes region do not induce large abrupt changes overall, this may not be the case in other regions. For example, abrupt changes are expected in tropical regions due to field preparation (i.e. forest clearing), or in (semi-)arid regions due to sudden increases in biomass productivity. However, in some cases it may be related to large-scale land dynamics other than LSAs, such as large smallholder deforestation fronts that may later be replaced by LSAs (Arvor et al., 2012; Graesser et al., 2018). There, the following spatial analysis will confirm or not the presence of LSAs based on the particular spatial arrangement and composition of this form of intensive agriculture.

## 5.1.2 How and Why?

### *5.1.2.1 The change metric RGB composite map: a useful visualization tool to gain insight into the main land dynamics at the national scale*

In Chapter 2, BFASTm-L2, a change detection algorithm was proposed to detect and map at the national scale of Senegal the main LULC changes that occurred during the monitoring period (2003-2018). This map, based on the magnitude of change, is presented in Chapter 3. While useful, this map alone is not sufficient to disentangle LSAI from other drivers of LULC change. In order to gain some insight into the drivers of LULC change behind the BFASTm-L2 detected changes, we have assumed here that different drivers of LULC change (climatic, natural, anthropogenic non-agricultural and agricultural) have different signatures in terms of time series types of change (i.e. abrupt, gradual or seasonal: amplitude, NOS and LOS). Therefore, **our main objective was to find a set of spectro-temporal metrics that allow**

**characterizing the detected change in terms of the types of change involved**, thus contributing to the *How*, and at the same time to the *Why* facets of change.

The existing literature provides limited insights into the predominant types of change induced by the main drivers. As discussed in Section 1.3.3, the distinction between different drivers of change often relies on correlations with meteorological data, or on third-party data which affects the transferability of the approach. In Chapter 3, we hypothesized that :

- Climate variability and natural changes mainly affect the amplitude of the time series through different meteorological events, resulting in small expected changes in NOS or LOS over a 3-year period (in the absence of extreme events). Gradual natural changes were expected to induce mostly gradual changes without significant changes in the seasonal component;
- Anthropogenic non-agricultural changes were mostly expected to either remove vegetation biomass (i.e. urbanization, mining), or reduce it without subsequently affecting the seasonality of the time series (infrastructure construction).
- Finally, LULC changes associated with large-scale agriculture are mostly expected to affect the seasonality component of the time series, but also, as noted in Box 5.1.1, to induce abrupt changes depending on the environmental conditions of implementation.

To test these hypotheses, we first assessed the types of change detected by BFASTm-L2 through visual analysis of selected time series representing different known types of change (punctual study cases) of MODIS NDVI and TRMM pluviometry. In addition, we identified three metrics to improve discrimination between different types of detected change: a time series shape dissimilarity metric, an NDVI post/pre-change ratio, and the BFASTm-L2 magnitude of change. These metrics were chosen for their complementary sensitivities to different types of change. A sensitivity analysis using a synthetic dataset of time series with different types and intensities of change showed that:

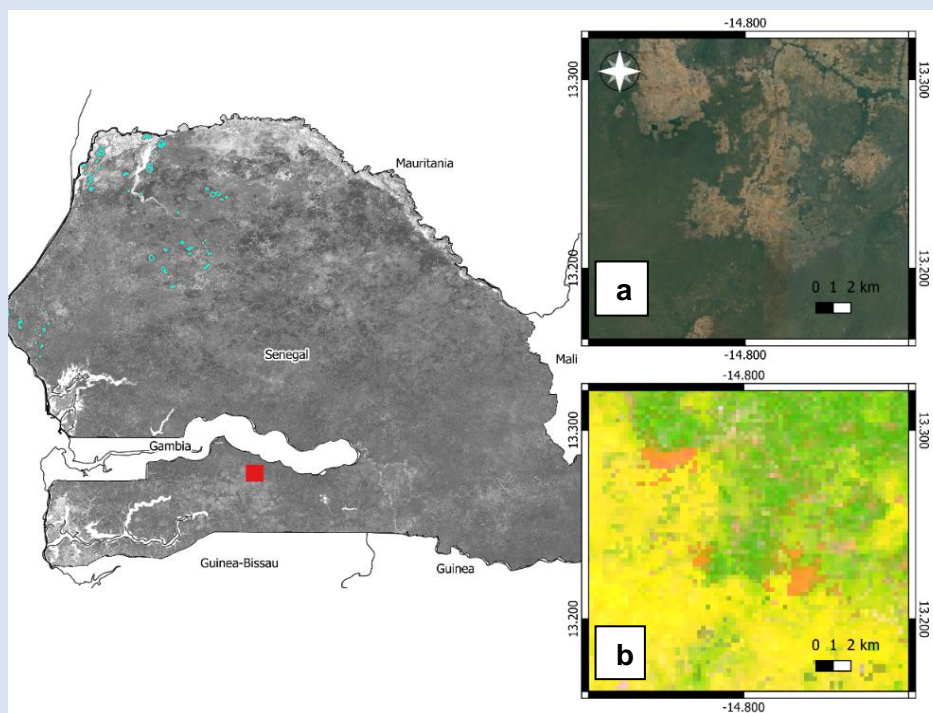
- The dissimilarity metric was highly sensitive to NOS and less sensitive to LOS changes;
- In contrast, the NDVI ratio showed high sensitivity to abrupt changes and, surprisingly, higher absolute values for LOS changes compared to amplitude or trend changes, and lower sensitivity to NOS. This metric also provided information on the direction of change, which helped to discriminate between agents of biomass removal;
- The magnitude of change, as introduced in Chapter 2, is known to be sensitive to abrupt and NOS changes.

As one can see, **these three metrics together allow to gain insight into the different types (and directions) of change involved behind a detected change**, and help to address **Research Question 2.1: Which time-series derived change metrics may be useful in distinguishing the different time-series types of change?**

In a subsequent step, these three metrics were combined into an RGB composite map, effectively highlighting different LULC changes. By considering the area and shape of the observed changes, it was possible to link different signatures of change (as indicated by the different colors of the RGB map) to different drivers of LULC change (i.e., climatic, natural, anthropogenic non-agricultural, and agricultural). In particular, these metrics, once combined and mapped, highlighted LSAI not only in the Niayes, but also in another ecoregion: the Senegal River valley. **This showed that LSAI-induced changes are detectable in different regions and share common characteristics, including higher dissimilarity and magnitude values compared to other land dynamics.** A caveat must be made here. While a mixture of seasonal and/or abrupt changes is often observed, in some rare cases neither is present. This is the case of the LSAs located in the Ferlo region and dedicated almost exclusively to the production of gum arabic. These are slow-growing plantations native to the Sahelian regions, planted over large areas that often appear sparse over time. They do not show significant seasonal variations during the first years of plantation. However, some of them show a positive NDVI ratio, but most of them are not intense and uniform enough across the field to be detected. Overall, the analysis of the RGB composite map performed in Chapter 3 allowed to address **Research Question 2.2: Are the main drivers of LULC change (i.e. climate variability, anthropogenic non-agricultural change, agricultural change, and natural change) more likely to cause a particular type of change in NDVI time series?**

**Box 5.1.2.– Composite RGB map: the case of the smallholder agriculture**

Despite the perception of limited change in the national composite RGB map (i.e., green pixels), it is important to note that much of the country's agriculture is small-scale and would be expected to produce seasonal changes due to land surface change in vegetation composition. Several factors contribute to the apparent underrepresentation of these changes on the map. First, the use of coarse spatial resolution data (250 m) focused on detecting large-scale agricultural activities, making it less likely to capture changes occurring at a scale smaller than the pixel size (~6 ha), such as those induced by numerous small fields contributing to a mixed signal. Second, the map statistics (for image stretching and visualization purposes) are calculated for the entire national territory, which may mask local variations unless they are large and intense. Finally, due to the low-intensity agricultural practices (with different crops, crop rotations) and the fact that in most cases the fields have been used for the same activity long before our monitoring period, the change associated with a change in practices, if detected, is of small magnitude. Exceptions may occur in cases of rapid change in contiguous areas, such as the observed agricultural expansion frontiers in forested regions (orange pixels in snapshot b)).



*Legend: Pioneer agricultural frontiers in the south of Senegal: a- Google Earth snapshot, b- RGB composite map snapshot based on BFASTm-L2 detected change in 2003-2018 MODIS NDVI SITS. Green pixels indicate small magnitude changes. Orange and yellow pixels indicate high magnitudes of change with either a decrease in NDVI average (orange pixels) or an increase in NDVI average (yellow pixels)*



### 5.1.2.2 *How well can LSAI be distinguished from other land dynamics?*

Chapter 3 provided valuable insights into the change signatures induced by major drivers of LULC change, using a combination of three spectro-temporal metrics. In particular, the RGB composite map showed similar change signatures (represented by pale pinks to whites, indicating higher dissimilarity and magnitude values) for most LSAs across two different ecoregions. This finding confirmed the viability of unsupervised LSAI detection. **The focus of the research then shifted to developing a method for automatically extracting and classifying hotspots of change potentially associated with LSAs through the integration of spatial analysis**, thus contributing to the *How* and *Why* facets of change. This was explored in Chapter 4, in two distinct steps.

**In the first step we sought to identify hotspots of change from a magnitude of change map** derived from BFASTm-L2 for subsequent analysis. This required an image in which LSAs stand out from their surroundings. Building on the results of Chapter 3, we introduced a change map derived from BFASTm-L2 change magnitude weighted by dissimilarity in Chapter 4. This weighting was intended to highlight pixels showing changes in the shape of the time series (i.e., NOS and LOS) (note: this metric was not used in Chapter 2 because the detection of abrupt changes was also desired in that step of the overall approach, see Box 5.1.1). Spatial analyses were then performed for two different ecoregions, the Niayes (semi-arid and subtropical dry) and the Senegal River (semi-arid), independently to account for potential masking of local variation by image preprocessing. The resulting weighted magnitude of change map significantly improved the spatial highlighting of LSAs, indicating that **LSAs in both regions have high weighted magnitudes that allow them to be distinguished from other land dynamics, particularly in the Niayes region**. From here, hotspots of change were then extracted from the weighted magnitude-based change map using contour analysis. While the spatial analysis was performed independently for each region, the contour analysis parameters were kept consistent to ensure the general applicability of the LSAI detection approach. The extracted hotspots showed spatial coherence. Depending on the region, about half (Niayes) to a quarter (Senegal River region) of the LSAs reported in the field database were considered "detectable" by this method, addressing **Research question 3.1: How can we automatically extract potential LSAI related hotspots of change?** There are three main reasons for this underreporting. First, the Senegal River region has a significant number of small LSAs, making their spatial identification difficult. Second, the presence of dynamic ecosystems, such as marsh-like wetlands and floodplains, in close proximity makes it difficult to detect changes in smaller objects. Finally, there are cases where changes in LSAs are not substantial enough-some dedicated to plantations and others with changes that do not persist over time (abandoned?) -resulting in a lack of significant magnitudes or dissimilarities.

**In the second step, we analyzed and classified the extracted hotspots, mainly to assess how well LSAs can be unsupervised distinguished from other land dynamics** and characterized them. This step mainly used evidence from spectro-temporal (from MODI NDVI SITS), structural and textural features computed from the change hotspots, applying unsupervised classification (K-means) to detect similar groups of objects. Structural and textural features were computed from more recent, higher spatial resolution Landsat imagery (from the end of the monitoring period) to assess post-change land use. Tests performed on the two different regions, show that structural features have varying discriminative power in function of the region analyzed. Therefore, structural features were excluded from the unsupervised classification analysis to maintain methodological robustness and genericity. In addition, it should be noted that not all LSAs have distinctive geometric shapes (e.g., see the example of large rubber plantations in Laos in Appendix B.1). On the other hand, since texture features (Haralick's features) were found to be not discriminative enough compared to other studies, they were also excluded from the unsupervised classification. Two main reasons were found to explain these results. First, the automatically extracted hotspot footprints often included external areas unrelated to the LSAs (i.e., areas that underwent changes likely related to field preparation, infrastructure construction, but also due to image preprocessing prior to segmentation), which significantly affected the calculated textures, especially when the LSAs were small. Second, the analyzed LSAs showed considerable spatial variability in terms of field plot shapes, sizes, and widths between plots, complicating the use of common textural features to identify LSAs.

Among the primary findings, we observed a distinct clustering of LSAs (composed of the three previously proposed spectro-temporal features and the Landsat-based NDBal, which allows the assessment of post-change land use), typically represented by 1-2 clusters within each region. However, the characteristics associated with LSAs are different for the two regions studied. While LSAs in the Niayes are mostly characterized by very large dissimilarities and high magnitudes of change, LSAs in the SR Valley are mostly characterized by relatively low magnitudes of change and dissimilarity (but very high NDVI ratios). This, in addition to answering **Research Question 3.2 (What spectro-temporal and spatial features are common to LSAs that may distinguish them from other land dynamics)**, confirms the need for analysis at the ecoregion level. Another important finding was that structural features (i.e., the proposed structural index related to the number of parallel lines, squares, angles, and circles found) to discriminate LSAs show promise and could be used to refine the identification of LSA-related clusters in specific regions. Finally, we found that the most closely related land dynamics, in terms of signatures of change, in both regions are those represented by marsh-type wetlands.

## 5.2 SYNTHESIS, RECOMMENDATIONS AND PERSPECTIVES

### 5.2.1 So, what do we detect?

In this thesis, we developed an algorithm (BFASTm-L2) that detects and selects the most significant (in terms of the importance of the change in the time series pattern/shape), persistent (evaluated over 3 years) LULC changes detected within long NDVI time series (here between 2003-2018). These are changes that are mostly associated with abrupt changes and/or changes in annual "peaks" (number of growing cycles). The coarse spatial resolution of the imagery used favors the detection of large scale changes (anthropogenic, climatic or natural).

Based on this change detection algorithm, two main products are proposed:

- The first is a **national-scale RGB composite map that highlights the most significant LULC changes with the greatest impact on land surface vegetation composition** (Chapter 3). By providing insight into the nature of the time series change behind the detected change, it allows for a comprehensive understanding of the most likely driver of change, an insight often missing from change maps based solely on magnitude of change. The fact that we do not use hard thresholds for any of the change metrics allows any change to be detected as long as it is contextually significant (with respect to its neighboring pixels). Large LULC changes such as those induced by climate variability (e.g. drought), natural changes (e.g. forest regeneration), anthropogenic non-agricultural activities (e.g. mining, infrastructure construction, urbanization), and agriculture (LSAI, but also smallholder pioneer fronts) can be visually detected on the map. Although the *When* facet of change was not explored here, it is worth mentioning that LULC changes observed within a single patch of change, may have been induced at different times. With respect to LSAs, these may be detected as long as they induced the most important (fast enough) LULC change within the monitoring period.
- The second output consists of **subnational maps showing the extracted hotspots of change** (as discussed in Chapter 4). This is an exploratory effort to automatically identify and locate potential LSAs. While user intervention is still required for cluster labeling, our results in Chapter 4 show that LSAs in each region studied fall predominantly into 1-2 clusters out of a total of 5-7. Furthermore, when all relevant LSA clusters are combined into a single cluster, more than half of the hotspots (65%-75%) are likely to be associated with LSAs. This has the potential to significantly streamline the process of identifying new LSAs, reducing the time and resources required to discover them. Notably, LSAs with a change area of less than 30 hectares

are not detected in this process, due to a size threshold applied to minimize the number of small hotspots detected. In addition, those that tend to cause slow trends or LOS changes (e.g. slow growing plantations) are not detected. In contrast, those that are homogeneous, producing monocultures with similar and uniform magnitude of change, are best detected. Since the changes often include external areas unrelated to the LSAs, and due to the convex hull transformation performed (aimed at avoiding complex shapes and obtaining compact polygons), field edge detections are not within the scope of this analysis.

### 5.2.2 Study limitations, recommendations and research perspectives

Chapter 3 defined the conditions for effective LSAI detection with BFASTm-L2: the analyzed region should ideally not include different ecoregions, the target change areas should be larger than 50 ha or far from unstable ecosystems, LSAs should involve fast-growing crops or important field preparation. In addition, when considering the application of this approach to alternative sensors than MODIS, it is important to keep in mind that the effectiveness of BFASTm-L2 relies on time series that must meet certain criteria: 1) have a high temporal frequency sufficient for accurate phenological representation, 2) have a minimum duration of 8 years, and 3) be both gapless and smoothed to minimize the occurrence of false detections. Currently, the application of BFASTm-L2 with Sentinel data is limited by its relatively short temporal depth. Conversely, in tropical conditions characterized by frequent cloud cover, the acquisition of high temporal frequency time series is challenging for both Sentinel and Landsat. In addition, the higher spatial resolution of these sensors would result in increased computational time, limiting the scalability of BFASTm-L2 for larger scale applications.

When exploring the different maps, and especially the RGB composite map, it is important to keep in mind that while certain change signatures appear to be associated with major drivers, there is no direct one-to-one relationship. A particular change signature could be attributed to different drivers of change, and conversely, a single driver could manifest itself in different change signatures, especially in different environments. Careful visual analysis should be performed using the size and geometry of the change patch, as well as contextual information (what do the changes look like in the region?). Finally, while still in an exploratory phase, it is recommended that the automated approach be applied to individual ecoregions. It is important to note that results may be different in regions where the diversity and size of agricultural systems makes LSAI less contrasted with other agricultural land uses (See the Argentine case in **Appendix B.2**).

Future research directions include: 1) improving the speed of BFASTm-L2 by using an implementation of BFAST monitor with automatic selection of stable historical periods, 2)

exploring and identifying spectro-temporal metrics that are responsive to LOS changes, 3) further research on the automatic extraction of structural features and how to integrate them into the classification analysis., and 4) exploring the ability of object-level spatio-temporal metrics to improve discrimination of LSAs. The rationale for the last proposal is based on the observation that the implementation and expansion of LSAs follows a systematic and sequential pattern in both time and space.

### 5.2.3 Operational perspectives

This work was undertaken with the specific aim of reducing the prevailing gaps in land information, particularly in countries of the Global South, in order to improve the transparency of land governance and to support informed decision-making on land issues. While the proposed approach does not provide a wall-to-wall area estimate of newly implemented large-scale agricultural investments (only significant land changes can be detected, which may be far below the contracted area), it goes beyond the mere quantitative analysis of land transactions by incorporating precise geo-location data. This advancement represents a critical step towards a comprehensive understanding of the full implications of large-scale agricultural investments, including social, environmental and economic impacts on a global scale.

However, as with other technology-based initiatives (land observatories, data-sharing platforms...), this approach alone is not sufficient to ensure that informed decisions are made (Gislain and Bourgoïn, 2023). As highlighted by other scholars (Bourgoïn et al., 2019; Özdoğan et al., 2018; Scoones et al., 2013), the proposition here is not to replace 'local voices' but rather to work alongside them. This is all the more important given the illustrative figures of Bourgoïn et al. (2019), who found that 78% of transactions reported in a participatory inventory were not present in previous inventories conducted by reference organizations such as Land Matrix and GRAIN. To facilitate the appropriation of the methodology, a Microsoft Planetary Computer API is currently being developed in Python language that will be hosted in Streamlit cloud. Along with this operational tool, there is a need for comprehensive training programs to effectively integrate it into existing practices and to ensure the transfer of skills and knowledge. It's important to note that confirmation of LSA detection is only possible after a well-organized field visit where ground truth information is carefully collected. This underscores the importance of practical, on-the-ground verification to validate the results of remote sensing and technological tools.

## REFERENCES

---

- Ajadi, O.A., Barr, J., Liang, S.-Z., Ferreira, R., Kumpatla, S.P., Patel, R., Swatantran, A., 2021. Large-scale crop type and crop area mapping across Brazil using synthetic aperture radar and optical imagery. *Int. J. Appl. Earth Obs. Geoinf.* 97, 102294.
- Anchang, J.Y., Prihodko, L., Kaptué, A.T., Ross, C.W., Ji, W., Kumar, S.S., Lind, B., Sarr, M.A., Diouf, A.A., Hanan, N.P., 2019. Trends in Woody and Herbaceous Vegetation in the Savannas of West Africa. *Remote Sens.* 11 (5), 576.
- Arvor, D., jonathan, M., Meirelles, M.S.P., Dubreuil, V., DURIEUX, L., 2011. Classification of MODIS EVI time series for crop mapping in the state of Mato Grosso, Brazil. *International Journal of Remote Sensing* 32 (22), 7847–7871.
- Arvor, D., Meirelles, M., Dubreuil, V., Bégué, A., Shimabukuro, Y.E., 2012. Analyzing the agricultural transition in Mato Grosso, Brazil, using satellite-derived indices. *Appl. Geogr.* 32 (2), 702–713.
- Awty-Carroll, K., 2019. Simulated NDVI time series repository. Open Science Framework. doi:10.17605/OSF.IO/TAF9Y. <https://osf.io/taf9y/>.
- Awty-Carroll, K., Bunting, P., Hardy, A., Bell, G., 2019. An Evaluation and Comparison of Four Dense Time Series Change Detection Methods Using Simulated Data. *Remote Sens.* 11 (23), 2779.
- Batterbury, S., & Ndi, F., 2018. Land-grabbing in Africa, in: , *The Routledge Handbook of African Development*. Routledge, pp. 573–582.
- Bégué, A., Arvor, D., Bellon, B., Betbeder, J., Abelleira, D. de, P. D. Ferraz, R., Lebourgeois, V., Lelong, C., Simões, M., R. Verón, S., 2018. Remote Sensing and Cropping Practices: A Review. *Remote Sensing* 10 (2), 99.
- Bellón, B., Bégué, A., Lo Seen, D., Almeida, C. de, Simões, M., 2017. A Remote Sensing Approach for Regional-Scale Mapping of Agricultural Land-Use Systems Based on NDVI Time Series. *Remote Sens.* 9 (6), 600.
- Bey, A., Jetimane, J., Lisboa, S.N., Ribeiro, N., Siteo, A., Meyfroidt, P., 2020. Mapping smallholder and large-scale cropland dynamics with a flexible classification system and pixel-based composites in an emerging frontier of Mozambique. *Remote Sens. Environ.* 239, 111611.
- Blaschke, T., 2010. Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing* 65 (1), 2–16.



- Borras, S.M., Franco, J.C., Moreda, T., Xu, Y., Bruna, N., Afewerk Demena, B., 2022. The value of so-called 'failed' large-scale land acquisitions. *Land Use Policy* 119, 106199.
- Bourgoin, J., Valette, E., Guillouet, S., Diop, D., Dia, D., 2019. Improving Transparency and Reliability of Tenure Information for Improved Land Governance in Senegal. *Land* 8 (3), 42.
- Brooks, E., Wynne, R.H., Thomas, V.A., Blinn, C.E., Coulston, J.W., 2014. On-the-Fly Massively Multitemporal Change Detection Using Statistical Quality Control Charts and Landsat Data. *IEEE Trans. Geosci. Remote Sensing* 52 (6), 3316–3332.
- Brooks, E., Yang, Z., Thomas, V., Wynne, R., 2017. Edyn: Dynamic Signaling of Changes to Forests Using Exponentially Weighted Moving Average Charts. *Forests* 8 (9), 304.
- Brown, J.C., Jepson, W.E., Kastens, J.H., Wardlow, B.D., Lomas, J.M., Price, K.P., 2007. Multitemporal, Moderate-Spatial-Resolution Remote Sensing of Modern Agricultural Production and Land Modification in the Brazilian Amazon. *GIScience remote sens.* 44 (2), 117–148.
- Browning, D.M., Maynard, J.J., Karl, J.W., Peters, D.C., 2017. Breaks in MODIS time series portend vegetation change: verification using long-term data in an arid grassland ecosystem. *Ecological applications : a publication of the Ecological Society of America* 27 (5), 1677–1693.
- Budde, M., Tappan, G., Rowland, J., Lewis, J., Tieszen, L., 2004. Assessing land cover performance in Senegal, West Africa using 1-km integrated NDVI and local variance analysis. *J. Arid Environ.* 59 (3), 481–498.
- Bullock, E.L., Woodcock, C.E., Holden, C.E., 2020. Improved change monitoring using an ensemble of time series algorithms. *Remote Sens. Environ.* 238, 111165.
- Chance, E., Cobourn, K., Thomas, V., 2018. Trend Detection for the Extent of Irrigated Agriculture in Idaho's Snake River Plain, 1984–2016. *Remote Sensing* 10 (1), 145.
- Chen, G., Liu, Z., Wen, Q., Tan, R., Wang, Y., Zhao, J., Feng, J., 2023a. Identification of Rubber Plantations in Southwestern China Based on Multi-Source Remote Sensing Data and Phenology Windows. *Remote Sens.* 15 (5), 1228.
- Chen, J., Jönsson, P., Tamura, M., Gu, Z., Matsushita, B., Eklundh, L., 2004. A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky–Golay filter. *Remote Sens. Environ.* 91 (3-4), 332–344.
- Chen, S., Olofsson, P., Saphangthong, T., Woodcock, C.E., 2023b. Monitoring shifting cultivation in Laos with Landsat time series. *Remote Sens. Environ.* 288, 113507.
- Chen, Y., Lu, D., Moran, E., Batistella, M., Dutra, L.V., Sanches, I.D., Da Silva, R.F.B., Huang, J., Luiz, A.J.B., Oliveira, M.A.F. de, 2018. Mapping croplands, cropping patterns, and crop types using MODIS time-series data. *Int. J. Appl. Earth Obs. Geoinf.* 69, 133–147.

- Chughtai, A.H., Abbasi, H., Karas, I.R., 2021. A review on change detection method and accuracy assessment for land use land cover. *Remote Sensing Applications: Society and Environment* 22, 100482.
- Cotula, L., 2012. The international political economy of the global land rush: A critical appraisal of trends, scale, geography and drivers. *The Journal of Peasant Studies* 39 (3-4), 649–680.
- Cox, E., 1927. A method of assigning numerical and percentage values to the degree of roundness of sand grains. *Journal of Paleontology* 1 (3), 179–183.
- Curtis, P.G., Slay, C.M., Harris, N.L., Tyukavina, A., Hansen, M.C., 2018. Classifying drivers of global forest loss. *Science (New York, N.Y.)* 361 (6407), 1108–1111.
- Davis, K.F., D'Odorico, P., Rulli, M.C., 2014. Land grabbing: a preliminary quantification of economic impacts on rural livelihoods. *Population and environment* 36 (2), 180–192.
- Degife, A.W., Zabel, F., Mauser, W., 2018. Assessing land use and land cover changes and agricultural farmland expansions in Gambella Region, Ethiopia, using Landsat 5 and Sentinel 2a multispectral data. *Heliyon* 4 (11), e00919.
- D'Odorico, P., Rulli, M.C., Dell'Angelo, J., Davis, K.F., 2017. New frontiers of land and water commodification: socio-environmental controversies of large-scale land acquisitions. *Land Degrad. Develop.* 28 (7), 2234–2244.
- Dutrieux, L.P., Verbesselt, J., Kooistra, L., Herold, M., 2015. Monitoring forest cover loss using multiple data streams, a case study of a tropical dry forest in Bolivia. *ISPRS Journal of Photogrammetry and Remote Sensing* 107, 112–125.
- Eckert, S., Kiteme, B., Njuguna, E., Zaehring, J., 2017. Agricultural Expansion and Intensification in the Foothills of Mount Kenya: A Landscape Perspective. *Remote Sens.* 9 (8), 784.
- Estel, S., Kuemmerle, T., Levers, C., Baumann, M., Hostert, P., 2016. Mapping cropland-use intensity across Europe using MODIS NDVI time series. *Environ. Res. Lett.* 11 (2), 24015.
- Evans, J., Geerken, R., 2004. Discrimination between climate and human-induced dryland degradation. *J. Arid Environ.* 57 (4), 535–554.
- Fan, H., Fu, X., Zhang, Z., Wu, Q., 2015. Phenology-Based Vegetation Index Differencing for Mapping of Rubber Plantations Using Landsat OLI Data. *Remote Sensing* 7 (5), 6041–6058.
- Faye, I.M., Benkahla, A., Touré, O., Seck, S., Ba, C.O., 2011. Les acquisitions de terres à grande échelle au Sénégal: description d'un nouveau phénomène, 46 pp.
- Gao, Y., Solórzano, J.V., Quevedo, A., Loya-Carrillo, J.O., 2021. How BFAST Trend and Seasonal Model Components Affect Disturbance Detection in Tropical Dry Forest and Temperate Forest. *Remote Sens.* 13 (11), 2033.
- Gellert Paris, R., Rienow, A., 2023. Using geospatial data to identify land grabbing. Detecting spatial reconfigurations during the implementation of the Nacala Development Corridor in

- Mozambique with remote sensing and land conflicts databases. *European Journal of Remote Sensing* 56 (1).
- Ghaderpour, E., Vujadinovic, T., 2020. Change Detection within Remotely Sensed Satellite Image Time Series via Spectral Analysis. *Remote Sens.* 12 (23), 4001.
- Giakoumoglou, N., 2021. PyFeats: Open-source software for image feature extraction. Zenodo.
- Gieseke, F., Rosca, S., Henriksen, T., Verbesselt, J., Oancea, C.E., 2020. Massively-Parallel Change Detection for Satellite Time Series Data with Missing Values, in: 2020 IEEE 36th International Conference on Data Engineering (ICDE). 2020 IEEE 36th International Conference on Data Engineering (ICDE), Dallas, TX, USA. 20/04/2020 - 24/04/2020. IEEE, pp. 385–396.
- Gower, J.C., 1975. Generalized procrustes analysis. *Psychometrika* 40 (1), 33–51.
- Graesser, J., Ramankutty, N., 2017. Detection of cropland field parcels from Landsat imagery. *Remote Sensing of Environment* 201, 165–180.
- Graesser, J., Ramankutty, N., Coomes, O.T., 2018. Increasing expansion of large-scale crop production onto deforested land in sub-Andean South America. *Environ. Res. Lett.* 13 (8), 84021.
- Grislain, Q., Bourgoïn, J., 2023. Land observatories, discourses and struggles beyond the smokescreen. A case study in Senegal. *Land Use Policy* 132, 106783.
- Grompone von Gioi, R., Jakubowicz, J., Morel, J.-M., Randall, G., 2012. LSD: a Line Segment Detector. *Image Processing On Line* 2, 35–55.
- Hamunyela, E., 2017. Space-time monitoring of tropical forest changes using observations from multiple satellites. Ph.D. Thesis, Wageningen University and Research, Laboratory of Geo-information Science and Remote Sensing, Wageningen, The Netherlands, 202 pp.
- Hamunyela, E., Rosca, S., Mirt, A., Engle, E., Herold, M., Gieseke, F., Verbesselt, J., 2020. Implementation of BFASTmonitor Algorithm on Google Earth Engine to Support Large-Area and Sub-Annual Change Monitoring Using Earth Observation Data. *Remote Sens.* 12 (18), 2953.
- Haralick, R.M., Shanmugam, K., Dinstein, I., 1973. Textural Features for Image Classification. *IEEE Trans. Syst., Man, Cybern.* SMC-3 (6), 610–621.
- Harding, A., Chamberlain, W., Anseeuw, W., Manco, G., Niassy, S., 2016. Large-scale land acquisitions profile. Senegal. [https://landmatrix.org/documents/60/LM\\_Country\\_Profile\\_Senegal\\_English.pdf](https://landmatrix.org/documents/60/LM_Country_Profile_Senegal_English.pdf). Accessed 1 June 2023.
- Hentze, K., 2019. How satellites can locate potential land grabs in Africa.

- Hentze, K., Thonfeld, F., Menz, G., 2017. Beyond trend analysis: How a modified breakpoint analysis enhances knowledge of agricultural production after Zimbabwe's fast track land reform. *Int. J. Appl. Earth Obs. Geoinf.* 62, 78–87.
- Hird, J.N., Castilla, G., McDermid, G.J., Bueno, I.T., 2016. A Simple Transformation for Visualizing Non-seasonal Landscape Change From Dense Time Series of Satellite Data. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sensing* 9 (8), 3372–3383.
- Horowitz, M.M., Salem-Murdock, M., 1993. Development-induced food insecurity in the middle Senegal Valley. *GeoJournal* 30 (2), 179–184.
- Hurni, K., Schneider, A., Heinimann, A., Nong, D., Fox, J., 2017. Mapping the Expansion of Boom Crops in Mainland Southeast Asia Using Dense Time Stacks of Landsat Data. *Remote Sens.* 9 (4), 320.
- Jamali, S., Jönsson, P., Eklundh, L., Ardö, J., Seaquist, J., 2015. Detecting changes in vegetation trends using time series segmentation. *Remote Sens. Environ.* 156, 182–195.
- Jenks, George F., Caspall, Fred C., 1971. Error on choroplethic maps: definition, measurement, reduction. *Annals of the Association of American Geographers* 61 (2), 217–244.
- Julien, Y., Sobrino, J.A., 2021. Introducing the Time Series Change Visualization and Interpretation (TSCVI) method for the interpretation of global NDVI changes. *Int. J. Appl. Earth Obs. Geoinf.* 96, 102268.
- Kennedy, R.E., Yang, Z., Cohen, W.B., 2010. Detecting trends in forest disturbance and recovery using yearly Landsat time series: 1. LandTrendr — Temporal segmentation algorithms. *Remote Sens. Environ.* 114 (12), 2897–2910.
- Kontgis, C., Schneider, A., Ozdogan, M., 2015. Mapping rice paddy extent and intensification in the Vietnamese Mekong River Delta with dense time stacks of Landsat data. *Remote Sensing of Environment* 169, 255–269.
- Kuemmerle, T., Erb, K., Meyfroidt, P., Müller, D., Verburg, P.H., Estel, S., Haberl, H., Hostert, P., Jepsen, M.R., Kastner, T., Levers, C., Lindner, M., Plutzer, C., Verkerk, P.J., van der Zanden, E.H., Reenberg, A., 2013. Challenges and opportunities in mapping land use intensity globally. *Current opinion in environmental sustainability* 5 (5), 484–493.
- Kuemmerle, T., Hostert, P., St-Louis, V., Radeloff, V.C., 2009. Using image texture to map farmland field size: a case study in Eastern Europe. *Journal of Land Use Science* 4 (1-2), 85–107.
- Kummerow, C., Barnes, W., Kozu, T., Shiue, J., Simpson, J., 1998. The Tropical Rainfall Measuring Mission (TRMM) Sensor Package. *J. Atmos. Oceanic Technol.* 15 (3), 809–817.
- Lay, J., Anseeuw, W., Eckert, S., Flachsbarth, I., Kubitzka, C., Nolte, K., Giger, M., 2021. Taking stock of the global land rush: Few development benefits, many human and environmental risks. *Analytical Report III*, 12 pp.

- Leenstra, M., Marcos, D., Bovolo, F., Tuia, D. Self-supervised pre-training enhances change detection in Sentinel-2 imagery. Part of the Lecture Notes in Computer Science book series (LNCS).
- Leroux, L., Bégué, A., Lo Seen, D., Jolivot, A., Kayitakire, F., 2017. Driving forces of recent vegetation changes in the Sahel: Lessons learned from regional and local level analyses. *Remote Sens. Environ.* 191, 38–54.
- Lhermitte, S., Verbesselt, J., Verstraeten, W.W., Coppin, P., 2011. A comparison of time series similarity measures for classification and change detection of ecosystem dynamics. *Remote Sens. Environ.* 115 (12), 3129–3152.
- Li, Z., Fox, J.M., 2011. Integrating Mahalanobis typicalities with a neural network for rubber distribution mapping. *Remote Sensing Letters* 2 (2), 157–166.
- Liu, C., Chen, Z., Shao, Y., Chen, J., Hasi, T., Pan, H., 2019. Research advances of SAR remote sensing for agriculture applications: A review. *Journal of Integrative Agriculture* 18 (3), 506–525.
- Lu, D., Mausel, P., Brondízio, E., Moran, E., 2004. Change detection techniques. *Int. J. Remote Sens.* 25 (12), 2365–2401.
- Mardian, J., Berg, A., Daneshfar, B., 2021. Evaluating the temporal accuracy of grassland to cropland change detection using multitemporal image analysis. *Remote Sens. Environ.* 255, 112292.
- Masilūnas, D., Tsendbazar, N.-E., Herold, M., Verbesselt, J., 2021. BFAST Lite: A Lightweight Break Detection Method for Time Series Analysis. *Remote Sens.* 13 (16), 3308.
- Matthieu Molinier, Jukka Miettinen, Dino Ienco, Shi Qiu, Zhe zhu, 2021. Optical Satellite Image Time Series Analysis for Environment Applications: From Classical Methods to Deep Learning and Beyond, in: , *Change Detection and Image Time Series Analysis 2*. John Wiley & Sons, Ltd, pp. 109–154.
- McGarigal, K., 2002. Landscape Pattern Metrics, in: El-Shaarawi, A.H., Piegorsch, W.W. (Eds.), *Encyclopedia of environmetrics*. Wiley, Chichester, New York.
- Meshkini, K., Bovolo, F., Bruzzone, L., 2021. An Unsupervised Change Detection Approach for Dense Satellite Image Time Series Using 3D CNN, in: *An Unsupervised Change Detection Approach for Dense Satellite Image Time Series Using 3D CNN. IGARSS 2021 - 2021 IEEE International Geoscience and Remote Sensing Symposium, Brussels, Belgium. 11/07/2021 - 16/07/2021. IEEE*, pp. 4336–4339.
- Molinier, M., Miettinen, J., Ienco, D., Qiu, S., Zhu, Z., 2021. Optical satellite image time series analysis for environment applications: From classical methods to deep learning and beyond. *Change Detection and Image Time Series Analysis 2: Supervised Methods*, 109–154.

- Ngadi Scarpetta, Y., Lebourgeois, V., Laques, A.-E., Dieye, M., Bégué, A., 2024. Insight into large-scale LULC changes and their drivers through breakpoint characterization – An application to Senegal. *Int. J. Appl. Earth Obs. Geoinf.*, 2024.
- Ngadi Scarpetta, Y., Lebourgeois, V., Laques, A.-E., Dieye, M., Bourgoïn, J., Bégué, A., 2023. BFASTm-L2, an unsupervised LULCC detection based on seasonal change detection – An application to large-scale land acquisitions in Senegal. *International Journal of Applied Earth Observation and Geoinformation* 121, 103379.
- Nolte, K., Chamberlain, W., Giger, M., 2016. *International Land Deals for Agriculture: fresh insights from the Land Matrix: Analytical Report II*. Bern Open Publ, Bern, 68 pp.
- Ochtyra, A., Marcinkowska-Ochtyra, A., Raczko, E., 2020. Threshold- and trend-based vegetation change monitoring algorithm based on the inter-annual multi-temporal normalized difference moisture index series: A case study of the Tatra Mountains. *Remote Sens. Environ.* 249, 112026.
- Ozdogan, M., Yang, Y., Allez, G., Cervantes, C., 2010. Remote Sensing of Irrigated Agriculture: Opportunities and Challenges. *Remote Sensing* 2 (9), 2274–2304.
- Özdoğan, M., Baird, I., Dwyer, M., 2018. The Role of Remote Sensing for Understanding Large-Scale Rubber Concession Expansion in Southern Laos. *Land* 7 (2), 55.
- Piou, C., Lebourgeois, V., Benahi, A.S., Bonnal, V., Jaavar, M.e.H., Lecoq, M., Vassal, J.-M., 2013a. Coupling historical prospection data and a remotely-sensed vegetation index for the preventative control of Desert locusts. *Appendix A. Basic and Applied Ecology* 14 (7), 593–604.
- Piou, C., Lebourgeois, V., Benahi, A.S., Bonnal, V., Jaavar, M.e.H., Lecoq, M., Vassal, J.-M., 2013b. Coupling historical prospection data and a remotely-sensed vegetation index for the preventative control of Desert locusts. *Basic Appl. Ecol.* 14 (7), 593–604.
- Radwan, T.M., Blackburn, G.A., Whyatt, J.D., Atkinson, P.M., 2021. Global land cover trajectories and transitions. *Scientific reports* 11 (1), 12814.
- Saxena, R., Watson, L.T., Wynne, R.H., Brooks, E.B., Thomas, V.A., Zhiqiang, Y., Kennedy, R.E., 2018. Towards a polyalgorithm for land use change detection. *ISPRS Journal of Photogrammetry and Remote Sensing* 144, 217–234.
- Schoneveld, G.C., 2011. The anatomy of large-scale farmland acquisitions in sub-Saharan Africa, <http://www.jstor.com/stable/resrep02317.1>, 32 pp.
- Schoneveld, G.C., 2014. The geographic and sectoral patterns of large-scale farmland investments in sub-Saharan Africa. *Food Policy* 48, 34–50.
- Scoones, I., Hall, R., Borrás, S.M., White, B., Wolford, W., 2013. The politics of evidence: methodologies for understanding the global land rush. *Journal of Peasant Studies* 40 (3), 469–483.



- Sedano, F., Molini, V., Azad, M., 2019. A Mapping Framework to Characterize Land Use in the Sudan-Sahel Region from Dense Stacks of Landsat Data. *Remote Sens.* 11 (6), 648.
- Setiawan, Y., Yoshino, K., 2012. Change detection in land-use and land-cover dynamics at a regional scale from MODIS time-series imagery. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.* I-7, 243–248.
- Setiawan, Y., Yoshino, K., 2014. Detecting land-use change from seasonal vegetation dynamics on regional scale with MODIS EVI 250-m time-series imagery. *Journal of Land Use Science* 9 (3), 304–330.
- Singh, A., 1989. Review Article Digital change detection techniques using remotely-sensed data. *Int. J. Remote Sens.* 10 (6), 989–1003.
- Solly, B., Charahabil, M., Dieye, E., Sy, O., Barry, B., Sagna B., Faye, C., 2020. Impacts of natural and anthropogenic factor on the woody flora of Haute-Casamance (South Senegal): from perception to reality. *J. Appl. Sci. Envir. Stud.* 3 (2), 117–131.
- Stefanski, J., Chaskovskyy, O., Waske, B., 2014. Mapping and monitoring of land use changes in post-Soviet western Ukraine using remote sensing data. *Appl. Geogr.* 55, 155–164.
- Sultan, B., Janicot, S., 2003. The West African Monsoon Dynamics. Part II: The “Preonset” and “Onset” of the Summer Monsoon. *J. Climate* 16 (21), 3407–3427.
- Sy, C.B., Cissé, E.H.T., Ba, S., 2015. Étude participative sur les acquisitions massives de terres agricoles en Afrique de l’Ouest et leur impact sur l’agriculture familiale et la sécurité alimentaire des populations locales: état des lieux. Cas du Sénégal, 21 pp.
- Tang, J., Arvor, D., Corpetti, T., Tang, P., 2021. Mapping Center Pivot Irrigation Systems in the Southern Amazon from Sentinel-2 Images. *Water* 13 (3), 298.
- Tappan, G., Sall, M., Wood, E., Cushing, M., 2004. Ecoregions and land cover trends in Senegal. *J. Arid Environ.* 59 (3), 427–462.
- Tuia, D., Roscher, R., Wegner, J.D., Jacobs, N., Zhu, X., Camps-Valls, G., 2021. Toward a Collective Agenda on AI for Earth Science Data Analysis. *IEEE Geosci. Remote Sens. Mag.* 9 (2), 88–104.
- United Nations Office for the Coordination of Humanitarian Affairs (OCHA). Sahel Crisis: 2011-2017. <https://reliefweb.int/disaster/ot-2011-000205-ner>. Accessed 17 July 2023.
- Verbesselt, J., Hyndman, R., Newnham, G., Culvenor, D., 2010a. Detecting trend and seasonal changes in satellite image time series. *Remote Sens. Environ.* 114 (1), 106–115.
- Verbesselt, J., Hyndman, R., Zeileis, A., Culvenor, D., 2010b. Phenological change detection while accounting for abrupt and gradual trends in satellite image time series. *Remote Sens. Environ.* 114 (12), 2970–2980.
- Verbesselt, J., Zeileis, A., Herold, M., 2012. Near real-time disturbance detection using satellite image time series. *Remote Sens. Environ.* 123, 98–108.

- Verburg, P.H., Alexander, P., Evans, T., Magliocca, N.R., Malek, Z., Da Rounsevell, M., van Vliet, J., 2019. Beyond land cover change: towards a new generation of land use models. *Current opinion in environmental sustainability* 38, 77–85.
- Verburg, P.H., van de Steeg, J., Veldkamp, A., Willemsen, L., 2009. From land cover change to land function dynamics: a major challenge to improve land characterization. *Journal of environmental management* 90 (3), 1327–1335.
- Vogels, M., Jong, S.M. de, Sterk, G., Addink, E.A., 2019. Mapping irrigated agriculture in complex landscapes using SPOT6 imagery and object-based image analysis – A case study in the Central Rift Valley, Ethiopia –. *Int. J. Appl. Earth Obs. Geoinf.* 75, 118–129.
- Wardlow, B.D., Egbert S.L., Kastens, J.H., 2007. Analysis of time-series MODIS 250 m vegetation index data for crop classification in the U.S. Central Great Plains. *Remote Sens. Environ.* 108 (3), 290–310.
- Weiss, M., Jacob, F., Duveiller, G., 2020. Remote sensing for agricultural applications: A meta-review. *Remote Sens. Environ.* 236, 111402.
- Wessels, K.J., van den Bergh, F., Scholes, R.J., 2012. Limits to detectability of land degradation by trend analysis of vegetation index data. *Remote Sensing of Environment* 125, 10–22.
- Winkler, K., Fuchs, R., Rounsevell, M., Herold, M., 2021. Global land use changes are four times greater than previously estimated. *Nature communications* 12 (1), 2501.
- Woodcock, C.E., Loveland, T.R., Herold, M., Bauer, M.E., 2020. Transitioning from change detection to monitoring with remote sensing: A paradigm shift. *Remote Sens. Environ.* 238, 111558.
- Xiao, C., Li, P., Feng, Z., Liu, Y., Zhang, X., 2020. Sentinel-2 red-edge spectral indices (RESI) suitability for mapping rubber boom in Luang Namtha Province, northern Lao PDR. *Int. J. Appl. Earth Obs. Geoinf.* 93, 102176.
- Xiao, J., Moody, A., 2005. Geographical distribution of global greening trends and their climatic correlates: 1982–1998. *International Journal of Remote Sensing* 26 (11), 2371–2390.
- Yan, L., Roy, D.P., 2014. Automated crop field extraction from multi-temporal Web Enabled Landsat Data. *Remote Sensing of Environment* 144, 42–64.
- Ye, S., Rogan, J., Sangermano, F., 2018. Monitoring rubber plantation expansion using Landsat data time series and a Shapelet-based approach. *ISPRS Journal of Photogrammetry and Remote Sensing* 136, 134–143.
- Ye, S., Rogan, J., Zhu, Z., Eastman, J.R., 2021. A near-real-time approach for monitoring forest disturbance using Landsat time series: stochastic continuous change detection. *Remote Sens. Environ.* 252, 112167.

- Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., Xu, H., Tan, W., Yang, Q., Wang, J., Gao, J., Zhang, L., 2020. Deep learning in environmental remote sensing: Achievements and challenges. *Remote Sens. Environ.* 241, 111716.
- Zhao, K., Wulder, M.A., Hu, T., Bright, R., Wu, Q., Qin, H., Li, Y., Toman, E., Mallick, B., Zhang, X., Brown, M., 2019. Detecting change-point, trend, and seasonality in satellite time series data to track abrupt changes and nonlinear dynamics: A Bayesian ensemble algorithm. *Remote Sens. Environ.* 232, 111181.
- Zhu, X.X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F., Fraundorfer, F., 2017. Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. *IEEE Geosci. Remote Sens. Mag.* 5 (4), 8–36.
- Zhu, Z., 2017. Change detection using landsat time series: A review of frequencies, preprocessing, algorithms, and applications. *ISPRS Journal of Photogrammetry and Remote Sensing* 130, 370–384.
- Zhu, Z., Qiu, S., Ye, S., 2022. Remote sensing of land change: A multifaceted perspective. *Remote Sens. Environ.* 282, 113266.
- Zhu, Z., Woodcock, C.E., 2014. Continuous change detection and classification of land cover using all available Landsat data. *Remote Sens. Environ.* 144, 152–171.
- Zhu, Z., Zhang, J., Yang, Z., Aljaddani, A.H., Cohen, W.B., Qiu, S., Zhou, C., 2020. Continuous monitoring of land disturbance based on Landsat time series. *Remote Sens. Environ.* 238, 111116.

## APPENDICES

### Appendix A : A sample of LSAI-related study cases

Study	Data spatial resolution	Region	Ancillary data (other than training data)	Detected LSAI	Method	Land unit	Assumptions	Period	Estimates
<b>Mapping of intensive agricultural systems</b>									
(Arvor et al., 2011)	250 m	Mato Grosso (906 000 km <sup>2</sup> ) (Brazil)		Double cropping systems involving soybean, maize and cotton.	2 ML-based classifications for agricultural masking and crop classification. + temporal segmentation	Pixel	Cropland can be discriminated from local vegetation based on EVI-time series derived statistical metrics. Size threshold: 25 ha.	2006-2007	16800 km <sup>2</sup> (out of 56000 km <sup>2</sup> of cropland) planted with two commercial crops
(Bellón et al., 2017)	250 m	Tocantins state (Brazil)	TerraClass (agricultural mask)	Two double cropping systems involving soybean, maize and rice.	PCA-based GEOBIA + rule-based and phenological pattern visual analysis classification	Object	Dominant cropland where pasture & rangeland (based on TerraClass product) < 30% of land unit.	2013-2014	11193 km <sup>2</sup> of dominant cropland
(Sedano et al., 2019)	30 m	Northern Nigeria (21 Landsat scenes)	DEM; HBASE; (expert knowledge)	Main agricultural systems including irrigated agriculture	Knowledge-based expert system	Pixel	Extensive knowledge on phenological cycles and separability of the relevant land surfaces	2015	Irrigation agriculture: 2.2% of the land; 14 % of the floodplains.

(Xiao et al., 2020)	20 m	Luang Namtha Province (9325 km <sup>2</sup> ) (Laos)		Rubber plantations	Algorithm based on deciduous features, differentiated with Red Edge Spectral Indices (RESI)	Pixel	Rubber trees may be differentiated from evergreen forests based on their deciduous characteristics in the dry season	2016-2018	771 km <sup>2</sup>
(Vogels et al., 2019)	6 m	Central Rift Valley (669 km <sup>2</sup> ) (Ethiopia)		Large-scale agriculture	GEOBIA + Random Forest	Object	Textural, shape and neighbour features are discriminative	Nov., Dec. 2013 and Feb. 2014.	120 km <sup>2</sup>
(Tang et al., 2021)	10m	Mato Grosso (Brazil)		Pivot irrigation systems	PVANET, GoogLeNet, Hough transform.	Pixel		June-August 2017	741 km <sup>2</sup> of irrigated area
<b>Supervised change detection approaches: Post-classification approaches</b>									
(Eckert et al., 2017)	30 m	Foothills of Mount Kenya (2491 km <sup>2</sup> ) (Kenya)		Large-scale irrigated agriculture	Random Forest	2 x 2 km	Irrigated agriculture with vegetation productivity during the dry season. Greenhouses and water bodies covered more than 3% of land unit	1987- 2016 1987- 2002 2002 -2016	3% of the study area in 2016
(Özdoğan et al., 2018)	30 m	Southern Laos		Large-scale rubber concessions	Decision tree C4.5 classifier	pixel		2004-2012 in 1 year intervals	300 km <sup>2</sup> planted (2004-2012)
(Bey et al., 2020)	30 m	Gurué district (5606 km <sup>2</sup> ) (Mozambique)		Large-scale agriculture	2 Random Forest + object-based size thresholding (> 5 ha).	Pixel	Textural features are discriminative due to mechanization.	2006; 2012; 2016	67 km <sup>2</sup> (1.2%)

(Chen et al., 2023a)	10,20,30 m	Xishuangbanna region (China)	Three LULC third-party products	Rubber	Multisource phenology characteristics + Random Forest	Pixel		2014;2016; 2018; 2020	2020: 4199 km <sup>2</sup>
<b>Supervised change detection approaches: Trajectory classification approaches</b>									
(Hurni et al., 2017)	30 m	Southeast Asia (7 Landsat tiles (3 in Laos))		Boom crops (rubber, cashew, eucalyptus, sugarcane and coffee)	SVM	Pixel		2000 - 2014	7360 km <sup>2</sup>
(Curtis et al., 2018)	30 m (based on forest cover change maps)	Global	Tree cover maps; population and wildfire databases	Commodities related to deforestation: agriculture (including plantations), mining and energy infrastructure	Decision-tree model	10 x 10 km	Permanent conversion of forest/shrubland to non-forest land uses	2001-2015	Southeast Asia and Latin America: > 60 % of commodity-driven deforestation
(Chen et al., 2023b)	30 m	Laos		New plantations following large forest disturbances	CCDC-SMA	Pixel	Plantations induce a time series pattern change. A threshold on magnitude is applied	1991-2020	969 km <sup>2</sup> (0.4% of deforestation)
<b>Unsupervised change detection approaches</b>									
(Yan and Roy, 2014)	30 m	150 x 150 km agricultural regions (Texas, California, South Dakota)	Prior extraction of a crop and a crop field edge probability map.	Large-scale agricultural crop fields	Active contour segmentation; Watershed algorithm; Geometric based algorithm	Object	Crops are assumed to be pixels with consistently (i.e. d weeks) high seasonal NDVI.	2006-2010	

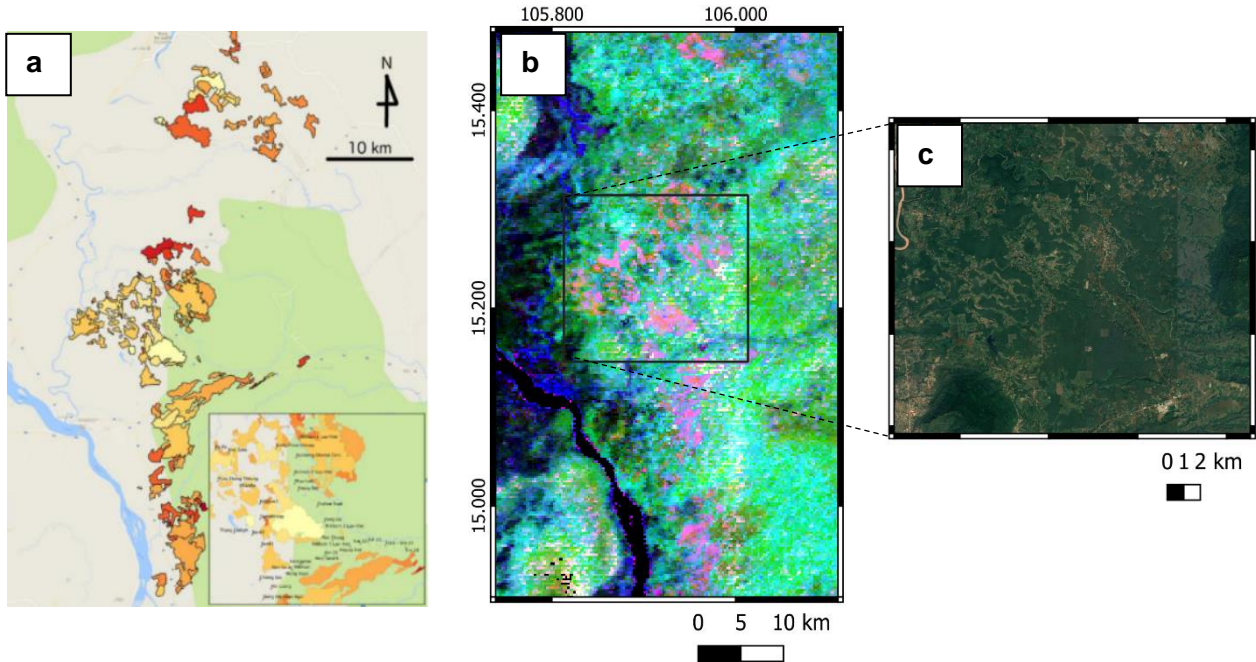
		(United States)							
(Kontgis et al., 2015)	30 m	Mekong River Delta (Vietnam)		Detection of single, double and tripple-cropped rice paddies	Pixel-based thresholds on EVI/NDWI and GEOBIA; Decision tree C4.5 classifier.	Pixel/object	Size threshold on land unit > ½ ha.	2000; 2010	% of rice farms with triple cropped fields increase from 34% (2000) to 62% (2010)
(Graesser et al., 2018)	30 m	Sub-Andean South America (48 % of continental area/ majority of region croplands)		Expansion of large-scale cropping systems on deforested land	Pixel-based random trees classification (at ecoregion-level) and object extraction based on histogram equalization (CLAHE) and adaptive threhsolding (Ath)	Pixel/object	Are only considered forest-to- cropland transitions; Size requirements: >50% of land object is estimated as cropland; Land object > 50 ha.	1990-2014 in 5 year intervals (timeframe of 1.5 year)	Acreage increase of 32% - 48%. Cumulative 1990-2014 cropland expansion is >300 900 km <sup>2</sup>
(Hentze, 2019)	250 m	Agro-ecological region IIa (AERIIa) (Zimbabwe)	SADC 1997 land cover map for crop masking; Dataset for agro-ecological zone stratification;	Changes in irrigation patterns	Seasonal Trend Analysis (STA) + BFAST	Pixel	Irrigation allows two growth cycles in subtropical Zimbabwe, allowing its differentiation from non-irrigated agriculture.	2000-2015	
(Ye et al., 2021)	30 m	Seima Protection Forest (11805 km <sup>2</sup> ) (Cambodia)	Forest map	Rubber plantations	Shapelet detection algorithm	Pixel	Rubber trees, compared to evergreen forests, have a time period of consistently low vegetation cover due to land clearing/ planting preparation	1995 - 2015	



(Gellert Paris and Rienow, 2023)	30 m, 250 m	Nacala Corridor (Mazambique)	Databases on land deals	Large-scale irrigated agriculture	1- LandTrendr; 2- statistical-based threshold	Pixel	Overlapping detected changes between 1 & 2 are related to land grabs; High productivity during the dry season	1994-2020	By district.
----------------------------------	-------------	------------------------------	-------------------------	-----------------------------------	---	-------	---	-----------	--------------

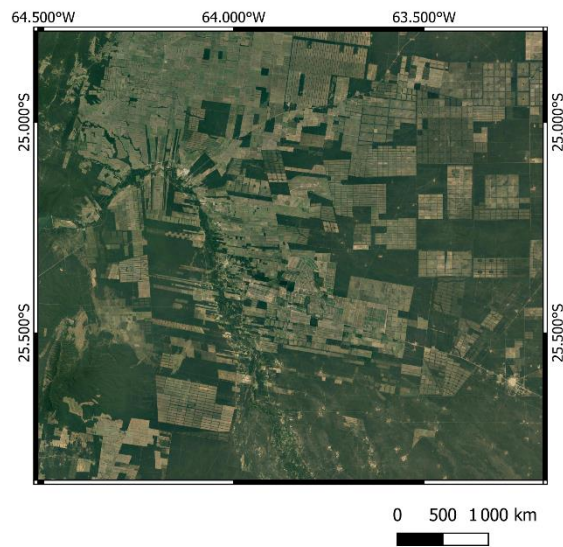
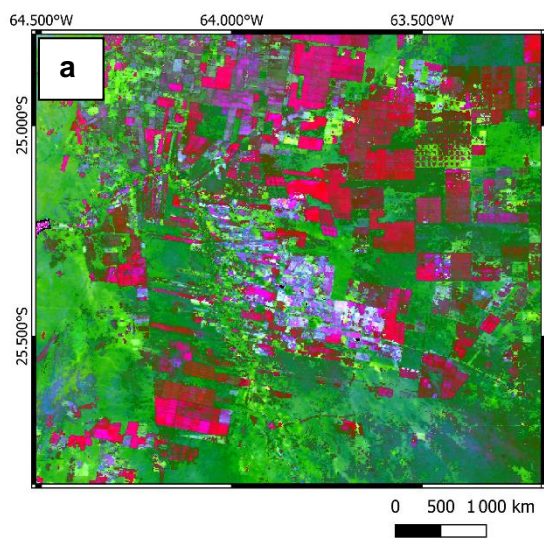
Appendix B : Case Studies from Laos and Argentina

B.1 : Southern Laos



Legend: a) Location and extent of rubber plantation in southern Laos (source: (Özdoğan et al., 2018) ) b) RGB composite map based on BFASTm-L2 detected change in 2003-2018 MODIS NDVI SITS c) Snapshot from Google Earth Imagery

**B.2 : Northern Argentina**



*Legend: Newly implemented LSAIs in northern Argentina. a) RGB composite map based on BFASTm-L2 detected change in 2003-2018 MODIS NDVI SITS*