Exploiting multi-year high-resolution Sentinel-2 image time series for mapping fallow practice in West Africa

Enzo Castro Alvarado

THÈSE POUR OBTENIR LE GRADE DE DOCTEUR DE L'INSTITUT DES SCIENCES ET INDUSTRIES DU VIVANT ET DE L'ENVIRONNEMENT - AGROPARISTECH

N° : 2023AGPT0015

En Géomatique

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

UMR TETIS

Potentialités des séries pluriannuelles Sentinel à haute résolution pour cartographier la pratique de la jachère en Afrique de l'Ouest

Présentée par Enzo Castro Alvarado
Le 19/12/2023

Sous la direction de Agnès BÉGUÉ
et le co-encadrement de Raffaele GAETANO et Louise LEROUX

Devant le jury composé de

Laurence Hubert-Moy, Professeure, Université Rennes 2	Rapporteur
Inbal Becker, Professeure, University of Maryland	Membre du jury
Thierry Bonaudo, Maître de conférences, AgroParisTech	Membre du jury
Bernard Tychon, Professeur, Université de Liège	Prèsident du jury & Rapporteur
Martin Brandt, Professeur assistant, Université de Copenhague	Invité

AgroParisTech

UNIVERSITÉ DE MONTPELLIER

Exploiting multi-year high-resolution Sentinel-2 image time series for mapping fallow practice in West Africa

Agroparistech, CIRAD

Enzo Castro Alvarado

19th December 2023

Abstract

Fallow mapping in West Africa is essential to accurately assess agricultural systems and its contribution to food security and agro-ecological sustainability of current practices, and yet the available mapping methodologies are not adapted to the environmental and cropping conditions encountered when addressing tropical smallholder agriculture. In this doctoral thesis, we explore different mapping strategies based on supervised classification techniques and making use of Sentinel-2 imagery and rainfall data as input, as well as multiple years of in-situ data to map fallow land at local scale in a Soudanian site in Burkina Faso (Koumbia) between the years 2016 and 2021. Results show that "traditional" machine learning based mapping approaches are not sufficient for detecting fallow land under the given pedoclimatic conditions, resulting in very low accuracy figures (e.g., F1-scores below the 0.2 mark). Most promising results were obtained when following a trajectory analysis approach, where a series of methodological adaptations had to be done to exploit annual data in a multi-year oriented manner. In this last case we reformulate the mapping problem to target non-active agricultural land (NAAL) as whole, obtaining F1-score ranging from 0.75 to 0.92 values when validating against complete (no data gaps) reference data set. Our results show that strategies that incorporate multiple years of spectral data in their learning process as a potential viable approach, where fallow land is not described by current status of land surface (i.e. land cover) but rather by the changes of it along the period that encircles the moment in which crop inactivity begins. However, results also indicate that the spatial application scope might be limited, with an augmentation of model uncertainty in areas where no ground truth data is available, highlighting the need to incorporate unsupervised approaches for enhanced extrapolation. On the other hand, more explicit multi-year strategies, where temporal analysis is delegated to model classifiers yielded marginally better results than annual direct mapping strategies, yet performances obtained do not reach satisfying results, with top average F1-score reaching the 0.44 mark. Methodological development is still required for both (a) exploiting more efficiently and direct manner multi-year data, and (b) building more cost-efficient unsupervised solutions that could be tested in areas with a reduce amount of ground truth data.

Keywords: Fallow, Africa, Sentinel, Burkina Faso, multi-year

Résumé

En Afrique de l'Ouest, la cartographie des jachères est essentielle pour caractériser les systèmes agricoles et évaluer avec précision la durabilité agroécologique des pratiques actuelles et leur contribution à la sécurité alimentaire. Cependant, les méthodes actuellement disponibles pour la cartographie des surfaces cultivées ne sont pas adaptées aux conditions environnementales et culturales rencontrées en contexte d'agriculture familiale en Afrique sub-saharienne. Dans cette thèse de Doctorat, plusieurs stratégies de cartographie des jachères basées sur des approches de classification supervisée ont été explorées. Pour ce faire, des séries temporelles d'images Sentinel-2 ainsi que des données pluviométriques ont été confrontées à une importante base de données collectées sur le terrain entre 2016 et 2021 sur le site de Koumbia, localisé en zone soudanienne au Burkina Faso. Les résultats de ce travail de thèse ont montré que les méthodes d'apprentissage automatique, dites "traditionnelles", ne permettent pas de détecter les surfaces en jachère dans les conditions pédoclimatiques de la zone d'étude. En effet, les précisions obtenues sont très faibles, avec des F1-scores inférieurs à 0.20. Les résultats les plus prometteurs ont été obtenus à partir d'une approche basée sur l'analyse des trajectoires. Une série d'adaptations méthodologiques a alors dû être effectuée afin de pouvoir exploiter les données annuelles dans leur contexte pluri-annuel. Ainsi, cette thèse s'est concentrée sur la classe "terres agricoles non actives" (NAAL en anglais) pour laquelle des valeurs de F1-scores entre 0.75 et 0.92 ont été obtenus, en considérant un jeu de données de référence complet. Par ailleurs, nos résultats ont mis en évidence que les stratégies de cartographie intégrant des informations spectrales pluri-annuelles dans le processus d'apprentissage constituent une approche viable, permettant de décrire les surfaces en jachère non pas par leur état actuel (c'est-à-dire par l'occupation du sol), mais par les changements au cours de la période qui jouxte la mise au repos des cultures. Cependant, nos résultats ont également montré que le domaine de validité spatiale de l'approche pouvait être limité, en raison de l'augmentation de l'incertitude du modèle dans les zones où aucune donnée de terrain n'est disponible. Cela souligne l'importance d'incorporer des approches non supervisées au processus de classification afin de permettre plus de fiabilité dans le processus d'extrapolation spatiale. Des stratégies pluri-annuelles plus explicites, où le processus d'analyse temporelle est délégué aux algorithmes de classification ont également été testées et ont montré des résultats légèrement améliorés par rapport aux stratégies de cartographie annuelles directes. Toutefois les performances obtenues restent modérées avec un F1-score moyen de 0.44. Des développements méthodologiques sont encore nécessaires

pour (a) exploiter de manière plus efficiente et directe les données pluri-annuelles, et (b) mettre en place des approches non-supervisées plus efficaces pouvant être testées dans les environnements pauvres en données de terrain.

Mots-clés: Jachère, Africa, Sentinel, Burkina Faso, pluriannuel

Acknowledgments

Enzo Castro Alvarado received for his work a fellowship from the French National Research Agency under the Investments for the Future Program ANR-16-CONV-0004 (DigitAg), and from CIRAD. The ground data collection was partly funded by CNES (APR TOSCA). The authors wish also to thank the European Space Agency (ESA) for the Sentinel 2 data, Theia pole for the calibration of the Sentinel-2 images, and Bertin Kaboré for collecting the 2019–2021 in-situ data.

Contents

Abbreviations

AAL Active Agricultural Land

AC Anti-causal

B2, B3, ... Band number 2, band number 3, ...

BRI Brightness Index

BI Brilliance Index

BI2 Brilliance Index II

CA Causal

CHIRPS Climate Hazards Group InfraRed Precipitation with Station data.

CI Color Index

CNN Convolutional Neural Network

FAO Food and Agriculture Organization of the United Nations

HRSL High Resolution Settlement Layer

iSDA Innovative Solutions for Decision Agriculture

JECAM Joint Experiment for Crop Assessment and Monitoring

LC Land Cover

LU Land Use

NAAL Non Active Agricultural Land

DAL Dynamic Agricultural Land

MSI Multi-Spectral Instrument

MNDWI Modified NDWI

NBR Normalized Burn Ratio

NDVI Normalized Difference Vegetation Index

NDWI Normalized Difference Water Index

NDRE Normalized Difference Red Edge Index

OA Overall Accuracy

OE Omission Error

OTB Orfeo ToolBox

PA Producer Accuracy

RI Redness Index

RF Random Forest

S2 Sentinel-2

SITS Satellite Image Time Series

SSA Sub-Saharan Africa

SWIR Short-Wave Infrared

SWNDVI Short-Wave NDVI

TCNN TempCNN

TempCNN Temporal Convolutional Neural Network

UA User Accuracy

VHSR Very Hight Spatial Resolution

VNIR Visible/Near infrared

List of Figures

List of Tables

Chapter 1

Introduction

1.1 General context

Sub-Saharan Africa (SSA) has been for decades the center of multiple initiatives, projects and research studies seeking ways to eradicate hunger, poverty and promote sustainable development in the subcontinent. In SSA over 22% of population is affected by chronic hunger, the highest proportion compared to other regions of the world (FAO, 2023). Climate change, demographic pressure and socio-political instability are some of the main drivers hindering development towards UN's 2030 agenda (Desa and others, 2016). West Africa in particular, has been under intensive demographic pressure, more than tripling its population since 1960s (FAO, 2022a) (see **Fig. 1.1**) and is expected to continue increasing to c.a. 1.5 billion by year 2100 (Vollset et al., 2020). Hence, an ever increasing strain in primary production systems is to be expected to cope with populations needs, specially in agriculture sector.

Sub-Saharan agriculture is composed by a large proportion of smallholders where small cropping fields are common, often less than 1 hectare defining a highly fragmented agricultural land, characterized by a high heterogeneity at the field level and where soil degradation, as well as a reduced and financially unaffordable access to mineral inputs and mechanization, limit crop production (Ruthenberg, 1971; Snapp et al., 2018; Tittonell and Giller, 2013). In this regard, Lowder et al. (2016) analyzed multiple data sources to quantify the type of farm, size and distribution of farmland worldwide, indicating that for the case of SSA, around 80% of farms are less than 1-2 ha and that these operate about 40% of farmland (see **Fig. 1.2**).

West African cropping systems have traditionally relied on long uncultivated periods of time to restore natural soil fertility, in some instances surpassing a decade between two periods of cultivation, after which the next cropping cycle is triggered by a removal and burning of all natural vegetation that has regrown during this idle period (slash-and-burn) (Ruthenberg, 1971; Manlay et al., 2000a; Samaké et al., 2005; Faye et al., 2021; Zoungrana, 1993). However, the necessity to enhance the agricultural production, driven by a growing population (Vollset et al., 2020), has promoted a shift in cropping systems,

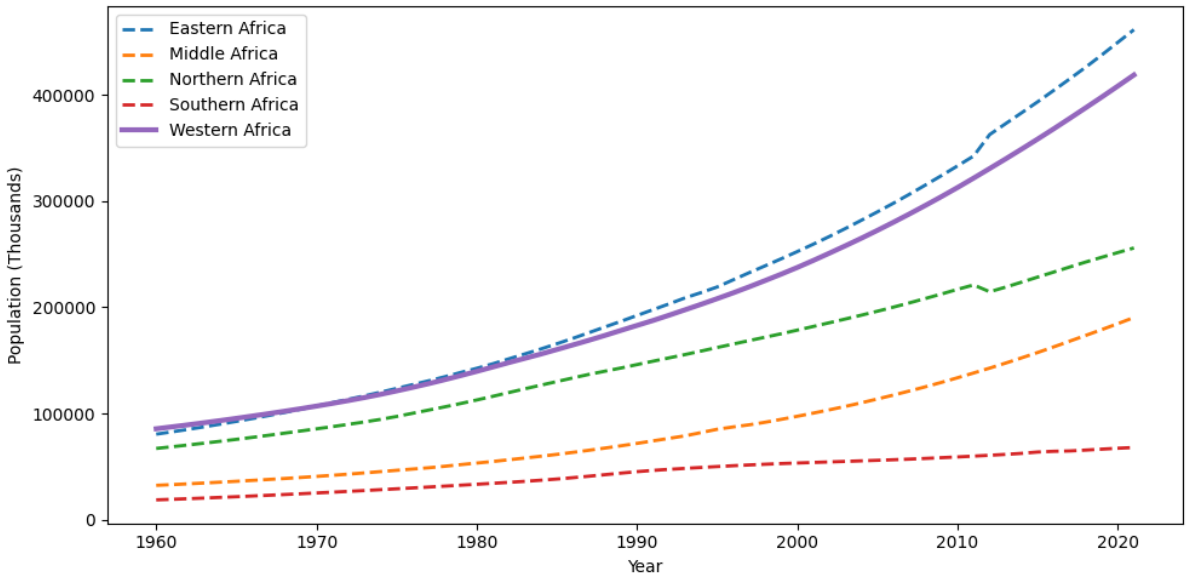


Figure 1.1: Annual population in Africa by regions (Source: FAO (2022a))

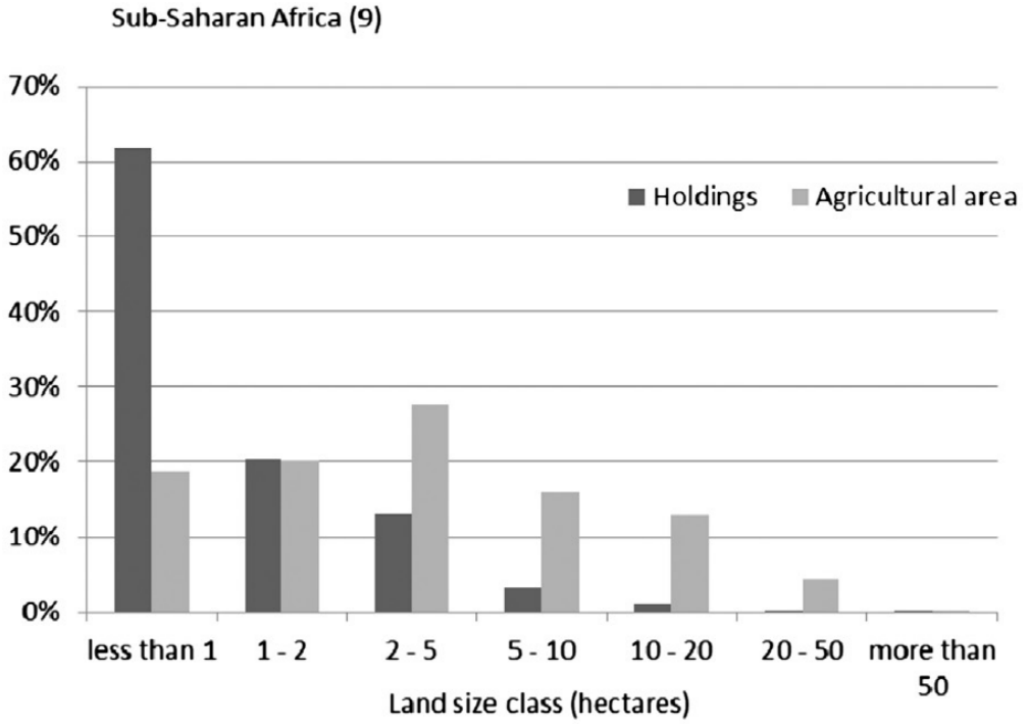


Figure 1.2: Sub-Saharan Africa average distribution of farms and farmland area by land size classes (Source: Lowder et al. (2016)). Number in parenthesis shows the number of countries considered.

which has been characterized by a reduction of fallow land and hence an increase of soil degradation. The quantification of both active (AAL) and non-active agricultural land (NAAL) is crucial not only for improving the estimation of potential production in the region but also for monitoring the sustainability of current intensification trends in local cropping practices from an agro-ecological perspective.

1.2 Fallow practice

Non-active agricultural land is often used as a synonym of "fallow", which according to FAO's definition accounts to "the cultivated land that is not seeded for one or more growing seasons. The maximum idle period is usually less than five years" (FAO, 2022b). Following this definition, what describes a fallow land is (a) the land has been previously cultivated, (b) the land is not currently being cultivated and (c) it is a temporary state that ends with a recultivation of land. This implies that the fallow land is actually defined by its past, current and future use rather than its sole current observable "land cover". However, more strict definitions also assume that (d) fallow fields exist as an agricultural practice within a cropping system as described in Bégué et al. (2018), which in turn implies that other abandoned agricultural land, due to reasons unrelated to cropping practices such as soil exhaustion, sociopolitical instability or simply manpower shortage should not be included within a strict sense "fallow" class. We could therefore establish that the idle period between two moments of observable active cropping is a necessary condition for validating a NAAL as "fallow" based exclusively on land cover data.

Fallow practice's role as mentioned before, relies essentially on its capacity of restoring soil fertility. Hence, *fallow* practice is an integral part in many cropping systems worldwide, but especially interesting in regions with low soil fertility and with reduced access to mineral fertilization, as it is the case in SSA. Nonetheless, in this same area fallows can also have additional functions beside soil fertility, as for example a way to provide animal feed or wood. Moreover, their importance has been further pointed out in recent times concerning the related implications on climate change mitigation through carbon sequestration and biodiversity conservation (Dayamba et al., 2016; Ringius, 2002). Shifting cultivation has been the primary agricultural system on SSA, which is characterized by a period of successive cultivation cycles 2-4 years/cycles, after which land would be abandoned once crop yields are no longer satisfactory. The period of abandonment (or fallowing) would vary depending on pedo-climatic conditions going from 2-3 years to more than 10 years in tropical regions. The duration of the active and inactive periods vary regionally, depending on the type of crops employed, pedo-climatic conditions and local socioeconomic factors (Ruthenberg, 1971) . During this uncultivated period, natural vegetation would gradually develop, recovering soil natural fertility until a new cultivation cycle is triggered by cutting out and burning all vegetation (commonly known as slash-and-burn) before cultivation. However, since the early 1960s scientific literature, authors have reported a shift in agricultural practices towards more intensive agricultural systems

in Africa, shortening farming cycles and exhausting soil fertility (Floret and Pontanier, 2001; Zoungrana, 1993; Jahel et al., 2018; de Ridder et al., 2004). This trend has been associated to demographic pressure and agricultural policies that encouraged specific, more intensive types of crops such as cotton in Burkina Faso (Jahel et al., 2018; Ouedraogo et al., 2010), characterized by the use of mineral inputs and mono-cropping (i.e. cultivate every season with same crop on same land). In fact, Quillemin (1956) already discussed the importance of *fallow* practice as well as the degradation of soil, and the consequent the loss of crop productivity, caused by demographic pressure and the shortening of crop cycles (see **Fig. 1.3**).

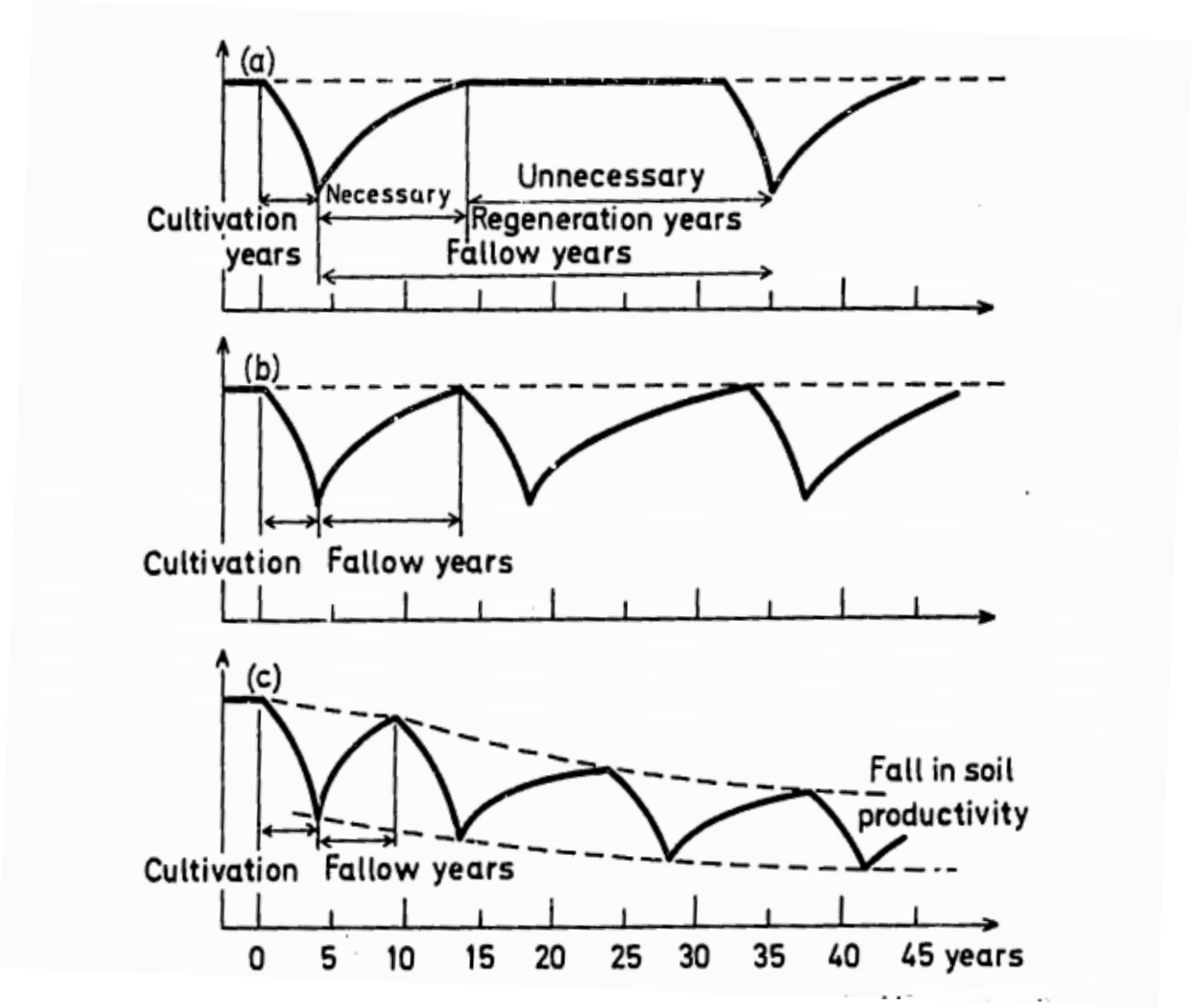


Figure 1.3: The relation between length of fallow and soil productivity in shifting cultivation. (a) Case where fallow period is longer than necessary for soil fertility recovery, (b) fallow period is sufficiently long to regenerate fertility and (c) fallow period is shorter than minimum needed for soil fertility regeneration (source: Quillemin (1956) and Ruthenberg (1971)).

Many are the ways and clues commonly used to characterize the implementation and effects of *fallow practices*. In most studies, the duration (or "age") of the idle period emerges as one of the main features used by authors to describe different categories of *fallow* land (Manlay et al., 2000b; Samaké et al., 2005; Faye et al., 2021; Zoungrana, 1993;

Akpo et al., 2002; Jouve, 1993). Such duration has an impact on the different stages of its "ecological" development, where the amount, type and predominance of vegetative species will vary, and hence in the plant/soil carbon storing capacity or potential nutritional levels for livestock feed (Manlay et al., 2000a; Dayamba et al., 2016; Akpo et al., 2002). Other abiotic parameters (temperature, soil type, humidity, etc) are also frequently taken into account, since they further condition its development, creating contrasting conditions among agro-climatic regions, yet *fallow* land is utilized for the same purposes. Moreover, we can identify more "pro-active" types of *fallow* where land is not simply left aside for a given period but actively managed, where we can find the so called *improved fallows* in which a selection and plantation of woody species is conducted to enhance fertility restoration during *fallow* period (Kaya and Nair, 2001). The human influence can also have an impact over the spatial arrangement of abandoned/fallowed fields as described by Ruthenberg (1971) in Tanzania, where a rotation system in which available land is divided and exploited during long 15-year cycles which include shorter (2-3 years) crop/fallow cycles before moving to a different area (Lericollais and Milleville, 1993). In Floret and Pontanier (2001) it is stated that for manual agricultural systems a higher probability of *fallowing* is observed in those fields further away from villages as they require more traveling time and workforce, whereas a higher cropping activity is more likely to be found in the vicinity of households.

Summing up, multiple factors can be identified which describe a widely heterogeneous typology of fallow land, which we tried to gather and illustrate schematically in **Fig. 1.4**. Most of the studies analyzing the role and potential benefits related to fallows in West Africa (and in entire Africa for that matter) are based theories and hypothesis built on a cohort of study sites, that provide insights on the factors that determine the characteristics of this practice. However, these studies depict a portrayal of fallow practices that dates back more than two decades and is likely in need of updating to reflect the integration of this practice into contemporary agronomic systems. Few studies have provided explicit and plausible estimates of the extent of this practice. Hence, considering the cumulative effect of decades of a growing population since the time when most of these studies were conducted and the changes that this implies in agronomic systems, there is a need for analyzing and quantifying its current state, especially in the Sudano-Sahelian area where most of smallholder agriculture traditionally relies on some sort of temporary agricultural land abandonment.

As a result of continued needs for accurate information regarding food production and available land to produce it, a growing interest in refining cropland mapping strategies has put the spotlight on the different uses of agricultural land worldwide with a particular attention to *fallow* practice given its rather extended use in agriculture sector and its clear impact over land mapping products accuracies.

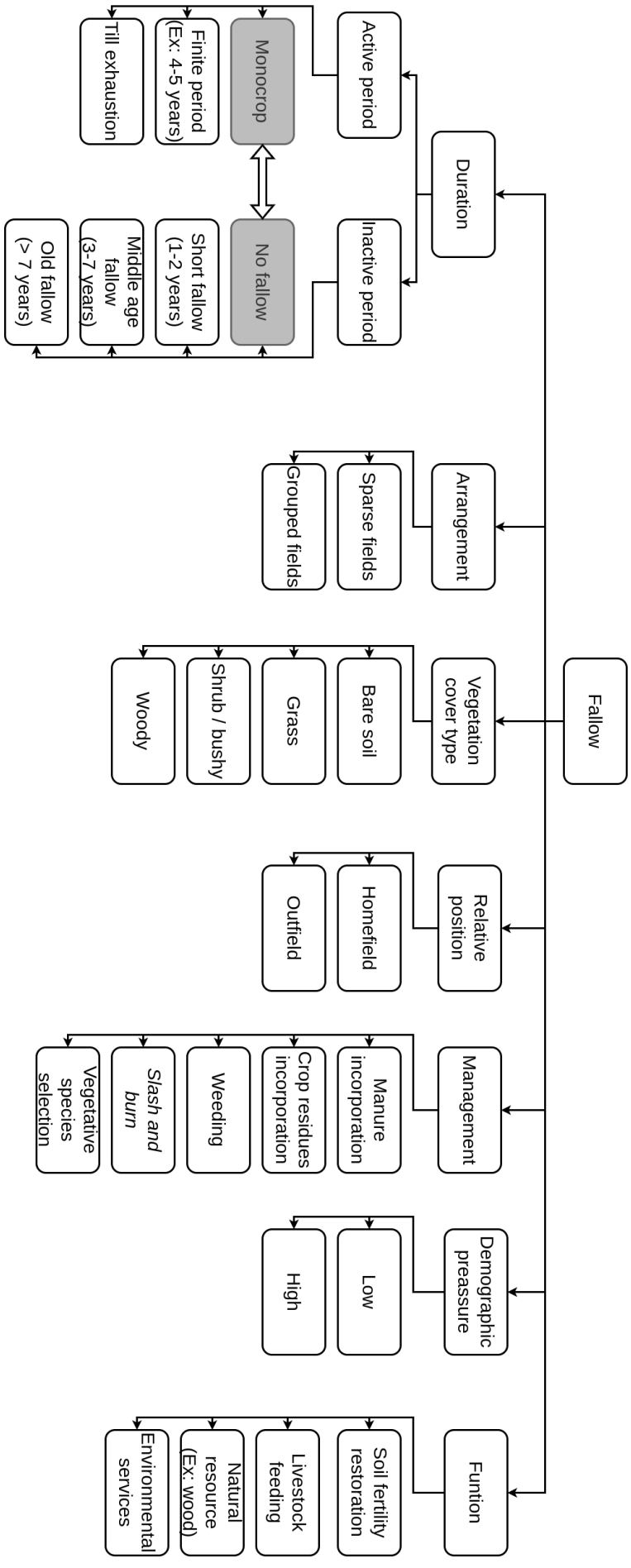


Figure 1.4: Scheme of the different characteristics that define fallow practice based on bibliographic review.

1.3 Fallow mapping studies

Satellite imagery has been proved to be a useful tool for generating land cover over broad regions such as *GlobCORINE* (Bontemps et al., 2009), *CGLS-LC100* (Buchhorn et al., 2020), or *Dynamic World* (Brown et al., 2022), allowing for a relatively economic solution and covering much larger areas compared to *in situ* surveys, and with continuous improvements in its spatio-temporal resolution. Nevertheless, NAAL mapping (or fallow land for that matter) has been, for the most part, disregarded by considering it an implicit part of a generic "cropland" class.

Indeed, few studies propose specific methodologies for mapping and quantifying non-active agricultural land through the analysis of satellite imagery. These works are quite varied in terms of the spatial scale, from regional level in California (Wallace et al., 2017) or in China (Zhang et al., 2014), to entire countries (Qiu et al., 2022), continental scale (Estel et al., 2015, 2016) or target specific agro-climatic zones such as the Sahel in Africa (Tong et al., 2020). Almost all these attempts for differentiating non-active agricultural land are based on the hypothesis that fallow land exhibits a significantly different spectral signature during the growing season with respect to active cropland, and then implement a classification strategy based on machine learning algorithms (Tong et al., 2020, 2022; Estel et al., 2015; Rufin et al., 2022) or a rule-based system at an annual scale (Wallace et al., 2017; Qiu et al., 2022; Zhang et al., 2014). To the best of our knowledge, only few studies have implemented strategies using input data which exceed a single growing season for determining whether a field is being actively cropped or not. In these cases (Yin et al., 2018; Dara et al., 2018), decades of Landsat data are used for determining a cropland probability, and land use classes are subsequently determined by inflection points of this probability through time. Moreover, in more recent studies (Rufin et al., 2022; Tong et al., 2022) a complementary verification is performed for "fallow" fields, requiring them to being cultivated once on previous years. A particularly interesting approach is proposed in Zhao et al. (2023) to estimate the proportion of *abandoned cropland* (hence not directly targeting *fallow* land) in Yunnan province in China using long time-series of Landsat imagery and a trajectory analysis based on long term land use annual mapping. Thus, establishing a multi-year temporal relationship as a strategy for NAAL identification, either for generating reliable reference data sets or as part of their mapping methodology, appears to be a point of convergence for scientific community. Regardless of the strategy implemented, these approaches often rely on the use of a cropland mask to focus the analysis on arable land only : *GlobCORINE* map cropland class in Estel et al. (2015, 2016), *USDA-CDL* in Wallace et al. (2017) and Wu et al. (2014), or *CGLS-LC100* in Tong et al. (2020). Moreover, data acquisition for training and/or validation is based for the most part on photo-interpretation of high resolution imagery or relying on the availability of a single season *in situ* data collection.

Summarizing, most of the remote sensing-based methods to monitor fallows require either (a) prior knowledge of cropland, (b) a crop/fallow reference data set obtained from

ground surveys or through photo-interpretation, and (c) contrasted seasonal spectral signatures among the land cover and land use (LCLU) classes. The first requirement usually implies the use of accessible global cropland products. In West Africa, the accuracy of such products is generally low to moderate, depending on the farming system in place (Leroux et al., 2014) with below 70% user accuracy on best cases (Samasse et al., 2018; Xu et al., 2019). The second requirement must cope with the lack of reliable ground truth data, which is often insufficient in both time and space, or available imagery does not allow the visual recognition of small and heterogeneous crop/fallow fields. Moreover, the latter requirement assumes a clear delimitation of cropland space (first requirement), however, under less favorable context, one can risk omitting *fallow* land as a consequence of a poor quality cropland mask. Alternatively, if no cropland mask is employed, the classification is further extended by including non-cropland vegetated spaces which are exposed to the same dynamics one can expect for *fallow* land (e.g. same vegetative species, abiotic conditions). This last issue connects directly with the third requirement (i.e. highly contrasted spectral signatures), which is rarely verified in tropical West Africa where, most of crops are *rainfed*, hence exhibiting any potential discriminating phenological development synchronously to other non actively managed vegetated spaces. Furthermore, given the rainfed nature of both cropped fields and naturally vegetated spaces, most of optical remote sensing imagery is unlikely to be fully exploitable due to cloud coverage.

1.4 Objectives

In light of the implications of *fallow* practice (or in a more broad sense, the presence of NAAL) in food security and other related environmental services, the evident lack of monitoring systems and the challenging "radiometric context" in West Africa, we state our main general objective for this thesis as:

> Explore and evaluate viable strategies for the identification of non active agricultural land (NAAL) in typical West African agro-systems through the analysis of high resolution satellite image time series relying on machine and deep learning classification algorithms.

We also establish a series secondary objectives for this thesis:

- Determine what are the properties and conditions observable through remote sensing data that allow *fallow* detection
- Quantify the need of reference data on land use/land cover which enable the different fallow mapping strategies (annual or multi-year).
- Provide recommendations for sampling protocols based on experience obtained in the different tested approaches.

To this end, we assume that recent satellite missions such as ESA's Sentinel-2, delivering

high-resolution and high frequency acquisitions in the optical multi-spectral domain over the globe, may indeed provide enough information to tackle some of these issues : (i) decametric spatial resolution can suitably address the spatial scale of the average field size in most of West African agrosystems, thus coping with landscape fragmentation issues, (ii) 5-day revisit time generally guarantees the acquisition of a sufficient number of images over a cropping season for tracking vegetation growth stages at field scale to some limited extent, and (iii) a sufficiently long archive of data (starting from late 2015) which largely encompass an entire *fallow* cycle (according to the FAO definition, see FAO (2022b)). Nonetheless, the diversity of cropping practices (sowing dates, cropping patterns, etc.) in conjunction with the local variability of the cropping conditions (soil, rainfall) may hinder the identification of specific agricultural land uses based on the sole image-derived spectral and temporal signatures (Inglada et al., 2015). We will focus on a representative study site of the Sudanian region in West Africa for which a ground truth dataset is available with observations of land use over multiple consecutive years covering the period 2013-2021.

The general approach followed in this thesis is based on first performing a series of preliminary analysis over available ground truth data in order to identify potential radiometric features that can be associated to non-active agricultural land (**Chapter 3**). Subsequently, we employ these insights to elaborate and test different classification strategies, separated into three different chapters (**Chapters 4**, **5** and **6**), each varying in technical complexity, in terms of training data set and relying on different manners to conceive non active agricultural land. In **Fig. 1.5** we illustrate a schematic summary of the structure of this work, where in **Chapter 2** (left side box) we describe the data and methods employed in the all the subsequent chapters. A series of preliminary analysis conducted over the ground truth data are presented **Chapter 3**, where we test different hypothesis and explore potential radiometric features that can describe *fallow* land. In **Chapter 4** we test two annual direct mapping strategies where we use traditional satellite image time series classification workflows for discriminating *fallow* land, basically providing a baseline approach that we intend to improve by shifting to multi-year analysis. In order to leverage the intrinsic multi-year nature of non-active agricultural land, in **Chapter 5** we present a first strategy based on a multi-year land cover/land use trajectory analysis for annual NAAL mapping. In **Chapter 6** we test two different strategies for mapping NAAL which, contrarily to our previous setting, advocates the use of multiple consecutive years of radiometric data as input data in order to build a proper NAAL classification model. Finally, in **Chapter 7** we provide an in-depth discussion about the outcomes of our experiments and we draw conclusions.

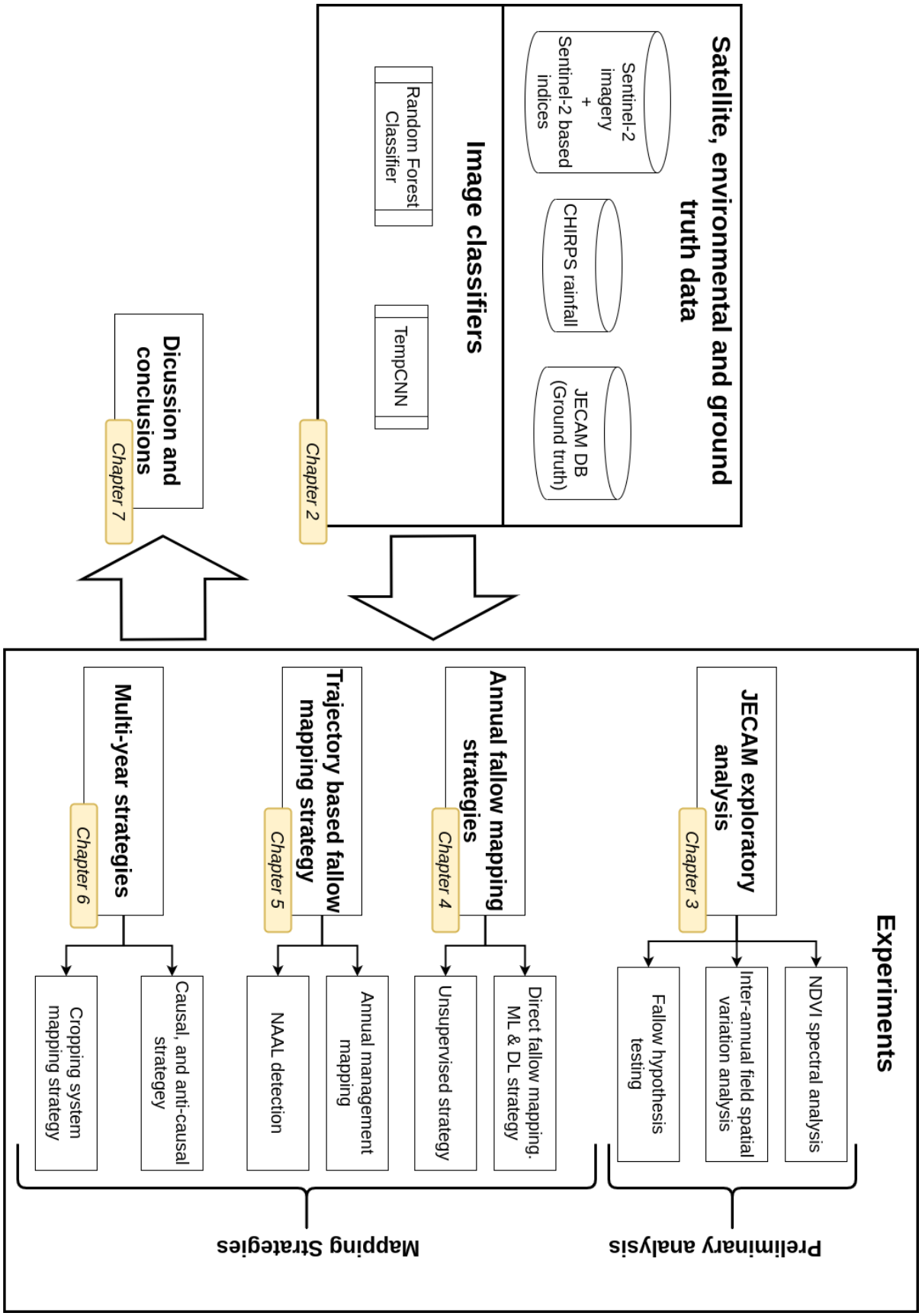


Figure 1.5: Schema presenting the general organization of the thesis.

Chapter 2

Material and methods

2.1 Study site

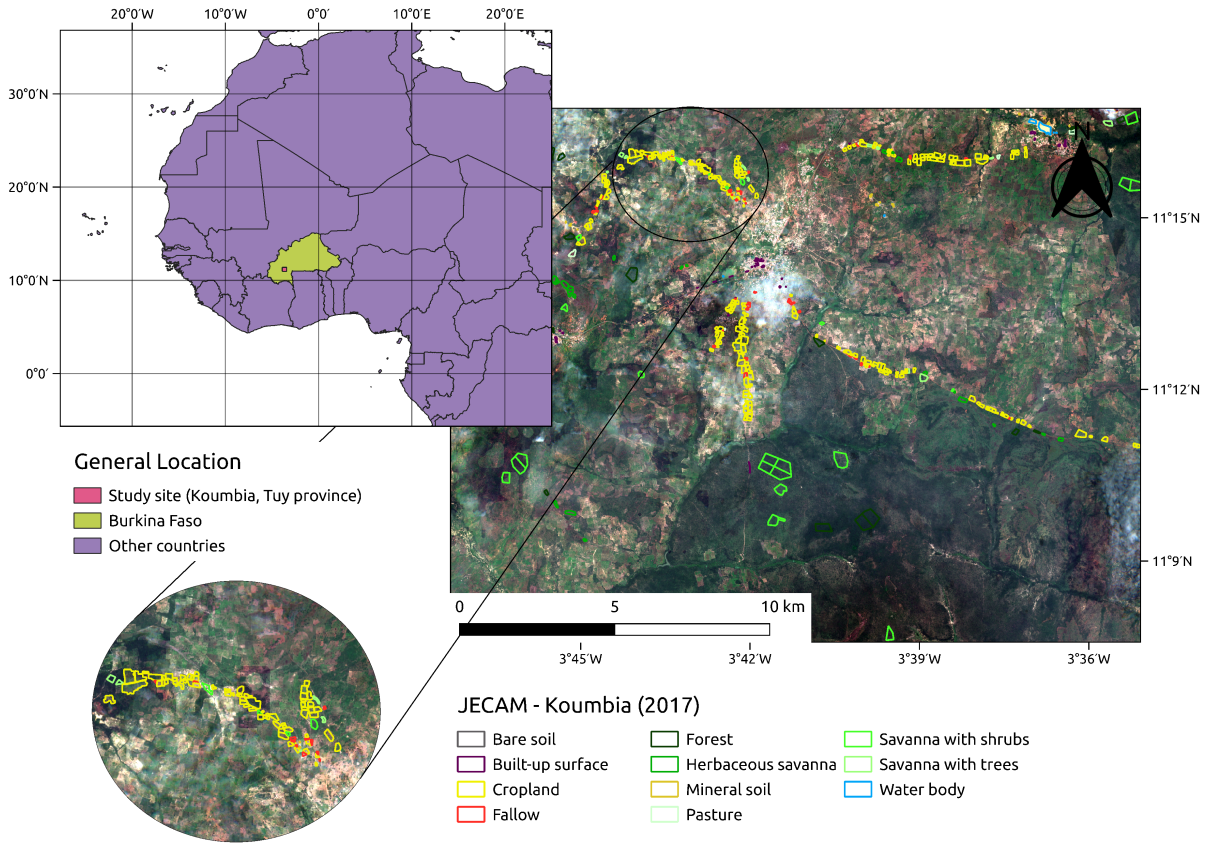


Figure 2.1: Location of the Koumbia study site in South-West of Burkina Faso (top left), JECAM ground data set polygons (Jolivot et al., 2021) and Sentinel-2 sub-image in 2017 (right), and a zoom-in (bottom left).

In this thesis work we focus on a study site located in the commune of Koumbia, Tuy Region, in South-Western Burkina Faso (see **Fig. 2.1**), covering an area of over 2500 km^2. In this area, rainfall distribution over a year is uni-modal and averaging between

900-1200 mm (see **Fig. 2.2**), which corresponds to typical values in the Sudanese agro-climatic zone (Abdoulaye et al., 2017). Cropping season begins on late April-May and extends till November-December, in concordance with the rainy season. As in most part of the Sahelian area, the study area is characterized by a tree-based cropping system, where multipurpose trees have been deliberately preserved and managed by farmers on agricultural land (Boffa, 1999). The cropping system is rainfed, relying mainly on maize-cotton rotations. Other crops include sorghum, millet, groudnuts, sesame (we group these groundnut and sesame within the *oilseed* label) and leguminous (e.g. cowpea) crops. Predominant soil textural types are sandy clay loam and sandy loam, accounting for 44% and 54% respectively in the region of study (Miller et al., 2021).

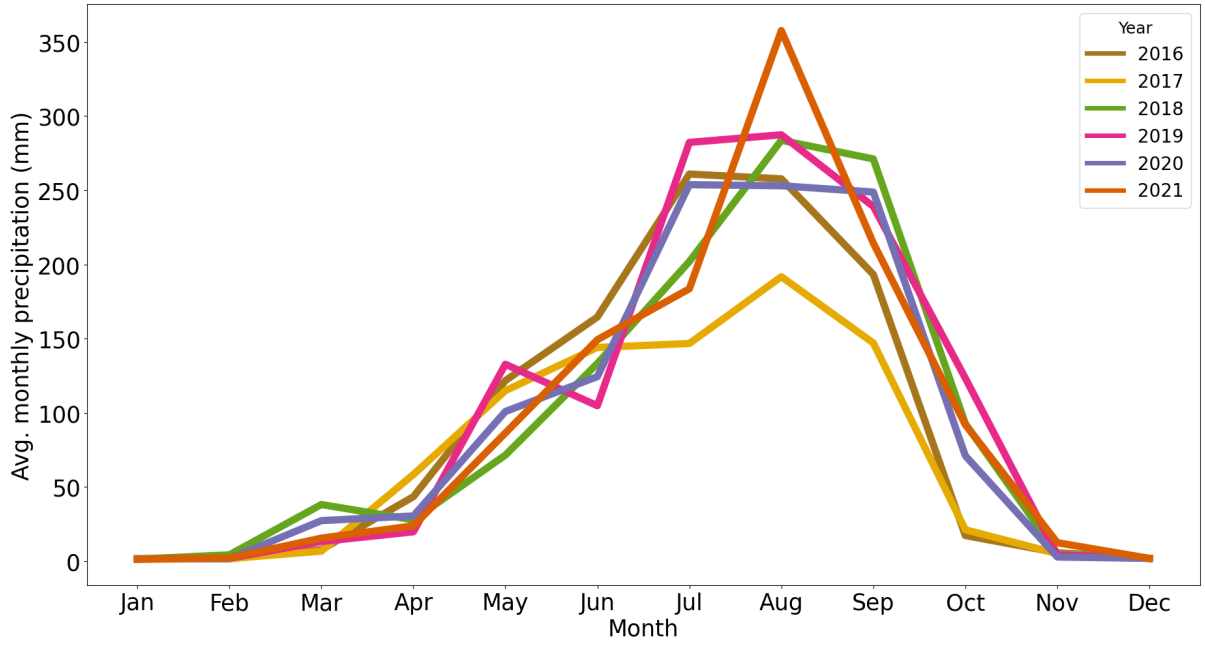


Figure 2.2: Monthly precipitation for all years considered (2016-2021) for Koumbia study site (Source: CHIRPS data Funk et al. (2014)).

2.2 Data

2.2.1 Reference data set

Across the multiple analysis performed, we employed as reference data the Koumbia subset of the recently published JECAM[1] (Joint Experiment for Crop Assessment and Monitoring) harmonized in-situ data set[2] (Jolivot et al., 2021). This ground data set includes land cover and agricultural land use in-situ observations collected annually over the Koumbia site during the 2013-2021 period (on average, c.a. 840 observations per year). The data set is built upon field missions conducted during the growing season where field

[1]https://www.jecam.org
[2]Access data at: https://doi.org/10.18167/DVN1/P7OLAP

operators collected waypoints for fields of at least 20 m^2 and were equipped with GPS tablet. Tablet disposed of a Very High Spatial Resolution (VHSR, SPOT 6/7 imagery) image acquisition of the site as a basemap, obtained during that same season. Field boundaries were digitized afterwards *ex situ* based on these same VHSR images. Each object (polygon) is annotated at both a *land cover* level, using a general nomenclature including the agricultural land in a single class (*cropland*, which includes also "fallow" land), and a set of classes related to natural areas (discriminating *herbaceous savanna* and *natural pastures* from *shrubby* and *woody savanna* and *forests*), *water* bodies and non-vegetated areas (*built-up surfaces* and *mineral soils*). In the JECAM data set, information on the crop type is also available for each polygon within the cropland class. We used this information to build an intermediate-level nomenclature named "Crop group", to keep *fallow* fields and active agricultural land (*cash crop*, *cereals*, *leguminous*, *oilseed*) separated. Details on the final nomenclature along with the surface covered by each class are given in **Tab. 2.1**.

Information for year 2019, originally missing from the published data set, was collected in 2020 by a local agent who carried out a delayed field survey, during which details on fallow land could not be collected.

Table 2.1: JECAM Koumbia data set area (ha) at *land cover* and *crop group* levels.

Land cover	Crop group	2015	2016	2017	2018	2019	2020	2021
Bare soil		1.60	1.41	0.66	1.33	0.00	3.40	1.90
Built-up surface		6.59	11.48	8.21	8.59	5.98	9.50	9.79
Herbaceous savanna		85.70	102.02	110.90	111.58	62.58	64.45	
Natural pastures		11.25	18.68	12.52	16.88	0.00	43.42	23.26
Shrubby savanna		214.43	229.51	241.94	231.96	214.43	219.04	235.59
Woody savanna		0.00	21.84	17.10	35.67	0.00	0.00	7.44
Forest		133.70	130.88	142.32	139.77	115.61	155.02	142.67
Water bodies		12.24	12.24	12.24	12.24	12.24	12.24	12.24
Cropland	**Cash**	73.71	105.78	212.01	94.22	102.59	81.16	114.78
	Cereals	171.53	179.51	232.07	159.50	161.14	109.74	172.20
	Leguminous	34.94	39.40	27.33	36.00	17.37	71.90	96.71
	Oilseed	50.04	42.60	35.22	15.63	8.52	28.69	34.08
	Fallow	32.13	20.64	21.79	7.69	0.00	8.74	9.41
Total		865.61	917.49	1077.30	872.65	700.92	809.10	924.28

2.2.2 Satellite and environmental data

In all the analysis conducted in the different chapters that compose this manuscript, we employed L2A Sentinel-2 (S2) image time series. The Copernicus Sentinel-2 mission includes a constellation of 2 satellites carrying a single multi-spectral instrument (MSI) with

13 spectral channels distributed in the visible/near infrared (VNIR) at 10 m spatial resolution, in short-wave infrared (SWIR) at 20 m spatial resolution, and in complementary bands at 60 m spatial resolution for cloud screening and atmospheric corrections (Drusch et al., 2012). All available images between 2016 and 2021 were downloaded from Theia[3] Land repository, and pre-processed using the open access MORINGA processing chain (Gaetano et al., 2019). Image pre-processing includes resampling the SWIR 20 m bands to 10 m resolution, and image time series gap filling of cloudy pixels using multi-temporal linear interpolation as explained in (Inglada et al., 2017). Gap-filled images were generated at a regular 10-day frequency and six vegetation-related radiometric indices (see **Tab. 2.2**) were calculated and appended to the reflectance bands for each time-step.

The soil type being a possible driver in the occurrence and duration of fallow practices (Samaké et al., 2005), we also computed for each year the median values of several soil indices using images acquired during the dry season (January - June) when the vegetation has not yet grown and the soil is visible. Four soil indices have been chosen, along with the Normalized Burn Area ratio as a variable describing the occurrence of fires, potentially related to slash-and-burn practices happening during the dry season prior to cultivation. A table resuming the computed indices used in this study, and their formula, is reported in **Tab. 2.2**.

Finally, in order to take into account rainfall in our analysis, we also acquired CHIRPS v 2.0 (Funk et al., 2014) monthly rainfall data, originally at 0.05° ($\approx$ 5.5 km) spatial resolution, reprojected over the S2 grid at 10 m resolution, using bicubic interpolation (see **Fig. 2.2**). Besides the fact that a significant downscaling error is likely, rainfall is known to be highly spatially variable in West Africa (Graef and Haigis, 2001; Lebel and Le Barbé, 1997), we make the assumption that these data may relate to the site-scale rainfalls.

In those instances where random forest-based classification tasks result into accuracy metrics beyond the 0.5 mark (see **Sec.??** an assessment of the contribution of each input variable on model decision is part of our analysis.

2.2.3 Complementary spatialized data

Additionally, as part of the preliminary analysis described in **Chapter 3** we download some complementary spatialized data which is not employed in classification strategies (i.e. **Chapters 4**, **5** or **6**). More specifically we retrieved the high resolution (30 meter pixel) ISDA top 20 cm (0-20 cm soil depth) USDA soil textural classes (Miller et al., 2021) data for our study site; We also retrieved the High Resolution Settlement Layer (HRSL) (Lab and for International Earth Science Information Network-CIESIN-Columbia University, 2016), containing information on human settlements in both urban and rural environments, as well as and population density for year 2015 at a pixel resolution of

[3]THEIA is the French Land Surface Pole gathering academic and public institutions to facilitate the use of Earth observation for monitoring continental surfaces. See `https://theia-land.fr` for more information.

Annual time-step	Index	Formula	Reference
10-day	Normalized Difference Vegetation Index	$NDVI = \frac{B8 - B4}{B8 + B4}$	Rouse et al. (1974)
	Normalized Difference Water Index	$NDWI = \frac{B3 - B8}{B3 + B8}$	McFeeters (1996)
	Brightness Index	$BRI = \sqrt{B2^2 + \cdots + B12^2}$	Inglada et al. (2017)
	Modified NDWI	$MNDWI = \frac{B3 - B11}{B3 + B11}$	Xu (2006)
	Short-wave NDVI	$SWNDVI = \frac{B11 - B8}{B11 + B8}$	Gao (1996)
	Normalized Difference Red Edge Index	$NDRE = \frac{B8 - B5}{B8 + B5}$	Barnes et al. (2000)
Seasonal (Jan-Jun)	Redness Index	$RI = \frac{B4^2}{B3^3}$	Mathieu et al. (1998)
	Color Index	$CI = \frac{B4 - B3}{B4 + B3}$	Mathieu et al. (1998)
	Brilliance Index	$BI = \sqrt{\frac{B3^2 + B4^2}{2}}$	Mathieu et al. (1998)
	Brilliance Index II	$BI2 = \sqrt{\frac{B3^2 + B4^2 + B8^2}{3}}$	Escadafal et al. (1989)
	Normalized Burn Ratio	$NBR = \frac{B8 - B11}{B8 + B11}$	García and Caselles (1991)

Table 2.2: List of the radiometric indices used. Provided bands refer to Sentinel-2 nomenclature ($B2$ is the *blue* band, $B3$ is *green*, $B4$ is *red*, $B5$, $B6$ and $B7$ are *red-edge* bands (700-800 nm),$B8$ is *near infrared* and $B9$, $B10$, $B11$ and $B12$ are *short-wave infrared* bands (950-2200 nm)). The first group corresponds to the multi-temporal indices computed at 10-day frequency over a whole year; The second group refers to the *soil* indices computed for the January-June period.

approximately 30 meter.

Notice that the ISDA soil data is a machine learning based predicted map which exploits over 40 000 soil samples in sub Saharan Africa. In this regard, we also generated a 10-meter resolution custom soil typology using Sentinel-2 imagery for enhanced resolution in our study site. For details in the protocol followed see annex section *Dry season soil map generation procedure*.

2.3 Image classification methods

Concerning the supervised classification strategies that were implied throughout the thesis, we mainly rely on two *state-of-the-art* machine learning methods which have proven their way in land cover/land use classification from satellite image time series. The first is the long time popular Random Forest technique, a classification method widely employed for highly multivariate data and consolidated for time-series classification tasks. As an alternative method, we considered the Temporal Convolutional Neural Network (TempCNN) (Pelletier et al., 2019) a novel technique from recent Deep Learning literature which, contrarily to Random Forest, explicitly models temporal correlation. These two methods (for which a brief description is given below) are used for implementing different NAAL discrimination strategies which vary in the amount of data given as input, nomenclature used in ground truth reference data (according to how we group the original JECAM land use classes) and the target class and period (usually a year).

For every classification task we provide a specific validation using dedicated portions of the JECAM land use/land cover data as reference data and providing the following confusion matrix derived accuracy metrics: overall accuracy (OA), user accuracy (UA), producer accuracy (PA) and per class harmonic mean (F1-score) (see Müller and Guido (2016)). Due to the highly unbalanced representation of classes, where cropland and natural spaces are highly over represented, OA is prone to be highly influenced by these majoritarian classes and hence not providing sufficient detailed insights. User accuracy and Producer accuracy can both provide better per-class details on commission and omission errors respectively. Yet, we deemed *F1-score* as a more representative metric of per-classes performance that can consolidate both UA and PA into a single metric with values ranging from 0 to +1 (with +1 being a perfect score). As we will specify in more details later in the manuscript (see **Sec. 2.3.3**), we will rely on a k-fold cross-validation (Müller and Guido, 2016), a process that splits reference data in k equally distributed subsets, of which k-1 concur to the model training and the remaining fold is used for testing. In principle, k different models are built at each training stage, and the final accuracy metrics are averaged over the k complementary models. Once classification models are trained and validated, we typically generate classification maps. These maps are constructed by first generating per model (5 maps) probability maps and then assigning the final class as the class with the highest average probability across all five probability maps.

2.3.1 Random Forest Classifier

Random Forest (Breiman, 2001) is an algorithm part of the ensemble decision tree family, commonly used in image classification tasks due to its predictive accuracy with largely multivariate data and ease of parameterization. This algorithm relies on a multitude of decision tree classifiers each performing a prediction based on a different random subsets of the input features provided. For its part, a decision tree classifier explore the feature space to find a suitable hierarchical partitioning of the source data based on simple feature-wise rules (e.g. thresholding), conceptually generating a tree representation of the input data whose leaves contain samples belonging to a single class (see **Fig. 2.3**). The main hyperparameters of a Random Forest classifier hence are the number of decision trees in the "forest", the maximum depth of these trees, the number of features to select for each tree and the level of "purity" of their leaves (i.e., the maximum number of samples in each leaf). Once the trees are built, the final prediction is the result of the most voted class across all trees that composes the forest. As a consequence, Random Forest predictions are based on statistically reliable decision rules, generally providing low levels of over-fitting. It is a very popular method in classical machine learning which is employed in a large number of different fields, including remote sensing where it has long been a *de-facto* standard for LULC mapping applications (e.g. Gbodjo et al. (2021); Inglada et al. (2017); Estel et al. (2015)). A complementary characteristic of Random Forest is that, once the model is trained, *feature importance* measures can be computed according to their occurrences in the different decision trees, and to the average level in which each tree considers them (features used closer to the tree root are generally more important). This allow for an in-depth inspection on the discriminative potential of each variable, and consequently, the possibility to optimize the model in terms of trade-off between speed and accuracy. However, with regards to its use in multivariate time-series, Random Forest exploits each feature independently, resulting in a lack of exploitation of any possible temporal correlation between timestep. For this reason, we decided to also test a second classification method in our experiments, which is introduced in the following subsection, especially conceived to explicitly model temporal correlations in time series of satellite images.

2.3.2 TempCNN classifier

The Temporal Convolutional Neuronal Network (TempCNN) is a deep learning classifier develop by Pelletier et al. (2019), specifically designed for working with time series of satellite images and explicitly modeling the temporal correlation between the different timesteps composing the series.

In a classical supervised deep neural network, data is processed through a series of consecutive layers which take as input data the outputs of the previous layer. Each layer is composed of a given number of units (or "neurons") and generates its outputs by applying a so called "activation" function to a linear combination of its inputs. There are

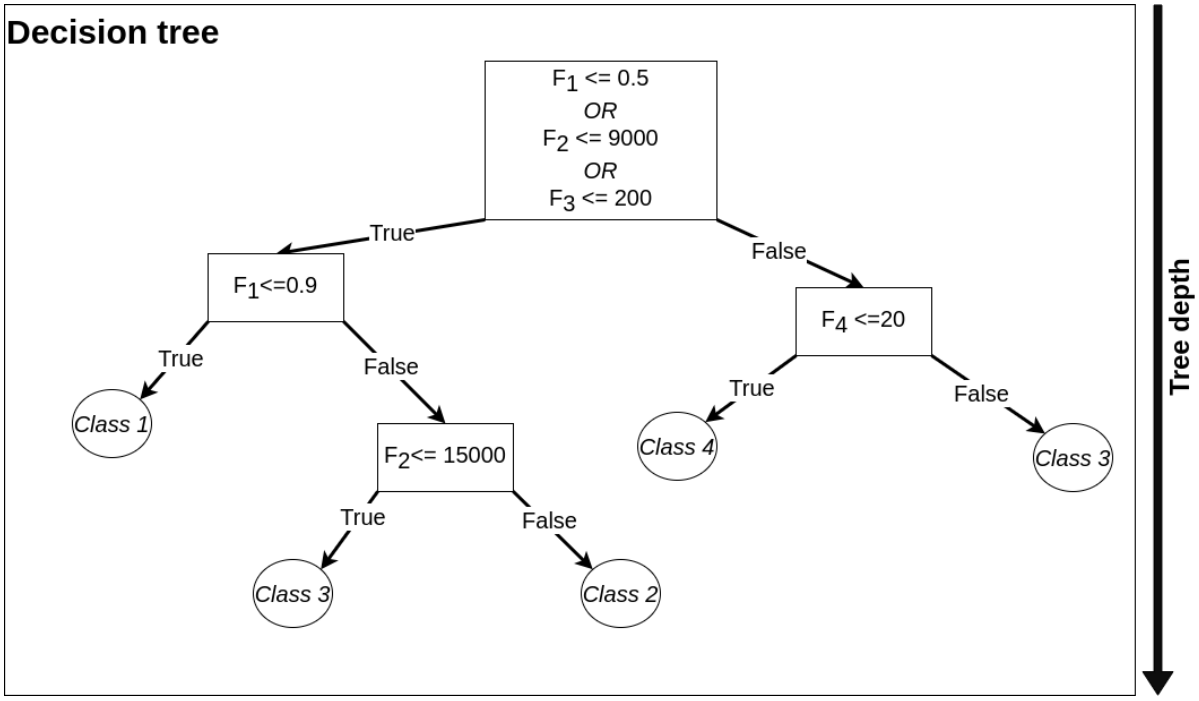


Figure 2.3: General schema of a decision tree in a four class classification task. F_1, F_2, F_3 and F_4 represent input features.

multiple types of layers that process data differently, where the most common types in a Convolutional Neuronal Network (CNN) are convolutional layers, non-linearity layers and pooling layers (Zhang et al., 2016). Training a deep learning model consists in finding the constants, or "weights", of each internal layer. This is usually done through an iterative process where the model is assessed through a *loss function* to minimize (e.g. the *categorical cross-entropy* (Goodfellow et al., 2016) for the typical multi-label classification task) and then weights adjusted performed through an *optimization* strategy. The optimization strategy (e.g. *stochastic gradient descent* (Bottou, 2010)) is what defines the actual learning process, basically performing a back-propagation of the current error by translating it into slight modifications to the layers' weights.

Back to TempCNN, The building blocks of the TempCNN architecture are one-dimensional convolutions operating in the time dimension and working jointly on all the set of input variables (see **Fig. 2.4**). The input time series flows through three consecutive convolution layers, followed by one dense layer that connects all neurons of the last convolution layer to each output unit, and finally a Softmax layer where outputs from the dense layer are employed to generate a probability-like classification with values ranging from 0 to 1, hence compose of C output units, where C is the number of target classes. Several additional mechanisms intervene in the learning strategy (*dropout*, *weight regularization*, *batch normalization*, see Pelletier et al. (2019) for details) to avoid over-fitting and make the architecture more robust and resilient to noisy data.

Notice that, in a classical deep learning classification task, data is usually divided into three independent parts *training* data set, used for learning patterns, a *test* data set for evaluating generated model accuracy and a *validation* data set use during the

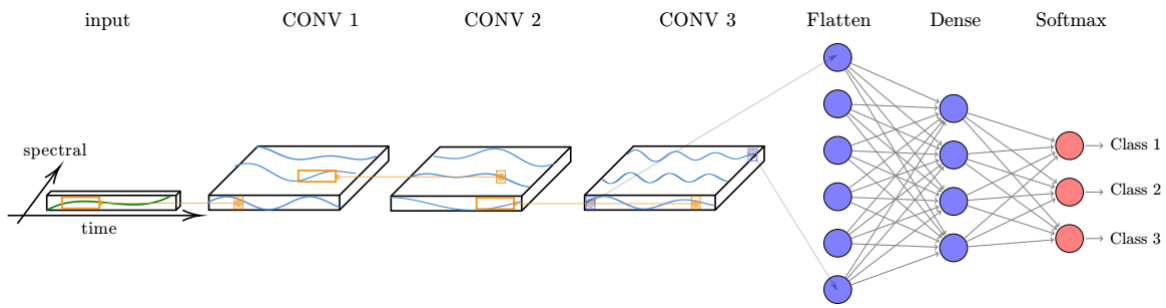


Figure 2.4: TempCNN architecture scheme. The network input is a multi-variate time series. Three convolutional filters are consecutively applied, then one dense layer, and finally the Softmax layer, that provides the predicting class distribution. (Source: Pelletier et al. (2019)).

training phase for evaluating and assessing model performance after each *epoch*, typically to stop the learning process when no significant improvements are measured on loss decay. However, since we work with a reduced data set, we opt for preserving data for model training and test only, using an heuristically determined number of epochs as learning duration. Details are given in the experimental sections.

2.3.3 General model training protocol

In this manuscript we will discuss multiple model training tasks that, unless specified otherwise, follow the same general protocol described in the following paragraphs. Due to an overall reduced amount of reference data (at least from the machine learning/deep learning point of view), especially when considering *fallow* land, it was decided to implement a k-fold cross validation design in the training procedure. Hence, data is split into five different (5 fold) subsets. In this manner, for each trial we train five models, each trained with a different combination of the available data. In each training job, model is trained with four out of the five folds (*k -1*) and validation is performed in the remaining fold. To cope with potential disparities in the amount of data for classes considered (i.e. some classes could not be equally represented among folds) and spatial auto-correlation issues in the validation process (Ploton et al., 2020), the sampling of data for the generation of folds happens at the level of polygons in the JECAM data base, following a stratified approach to ensure that the percentage of selected polygons is preserved in each class to the extent possible. We hence make the underlying assumption that, within each class, polygons have comparable areas (same order of magnitude), i.e. the distribution of samples across folds at pixel level will not differ significantly.

2.3.4 Classifiers setup

In this paragraph the basic parameterization for the two classification techniques described above are resumed. With due exceptions, these are the setting used in all the experiments.

Random Forest

Concerning the Random Forest (RF) classifier, we use the python package *scikit-learn* v1.2.2 implementation (Pedregosa et al., 2011). For all our experiences we set the parameterization based on previous ones over data in similar agro-climatic contexts and for comparable classification complexity. As we described in **Sec. 2.3**, RF has a reduced number of critical parameters to calibrate which may significantly impact the classification performances, which made the validation of the parameterization on our data sets relatively easy. More in detail, in our training setup we set the number of trees in forest to 400. Increasing this value does not generally translate into improved accuracies while requiring higher processing time, whereas the opposite (lower amount of trees) provides significantly lower performances. We also set the minimum number of samples required to split an internal node to 25 (with no limit to the depth of the trees), which favors generalization and speeds up processing especially at inference time. All the rest of parameters are left with default values, where the most relevant one is the maximum number of features employed at each tree split which is set by default to the root square of the maximum amount of input features.

As we will see in the different strategies tested further in this manuscript, avoiding unnecessarily complex models is fundamental for both obtaining sufficiently robust performances and optimizing computing resources.

A separated parameterization was employed for the case of the unsupervised approach detailed in **Chapter 4** where we used the Random Forest implementation of OrfeoToolbox (OTB) v7.2.0. (Inglada and Christophe, 2009). This setup used a maximum number of trees of 200, a training/validation ratio of 0 (i.e. all data is used for training) and default values for the remaining parameters. The latter setup is employed for an unsupervised approach whose accuracy was tested with ground truth data.

TempCNN

With regards to TempCNN, we only slightly modified the implementation provided by the original authors (see Pelletier et al. (2019)) in the *Keras* library (Chollet et al., 2015) with *Tensorflow* deep learning framework as backend (Abadi et al., 2016), and given the similar application context and volumes of data concerned, we also inherited the default parameterization of the original work, namely 64 units for each of the internal convolutional neural network layer, and 256 units for the subsequent dense layer. The number of input variables accepted by the model was modified accordingly to work with 16 variables per date (only decametrics variables from Sentinel-2 reflectances and indices). Moreover, *early stopping* of the learning procedure was deactivated and the number of training epochs is fixed at 20: the reason of this choice is that early stopping needs a dedicated portion of the reference data set (usually called *validation* data set in the deep learning context), which is not provided in our case in order to keep most of the reference data for the more critical training and test phases. We still keep 5% of training data

in each fold for model assessment after each training epoch (too few to decide for early stopping) in order to track the loss decay and check for the appropriateness of the training length. As in the original work, *dropout* rate is fixed to of 0.5 to minimize over-fitting issues.

In order to enable the fusion of per-fold model predictions, we modified the final layer in the architecture (the *softmax* layer) to directly yield per-class probabilities at pixel level instead of providing labelled samples. Final labelled maps are generated using these probability maps as described in **Sec. 2.3**.

Chapter 3

Exploratory analysis

In this chapter we will provide a deep insight into the JECAM data set and analyse the information it provides about the *fallow* practice. More specifically, we will try to understand how this practice relates to the local pedo-climatic conditions, and mine into the associated spectral, temporal and spatial features to better identify the methodological clues allowing for the mapping of this practice. We employ Burkina Faso JECAM's data subset (from now on also addressed as *Koumbia dataset*) as a spatial reference and Sentinel-2 images as input data to calculate spectral features, pre-processed as described in **Chapter 2**. Our idea is to formulate and test different hypotheses, as those introduced in **Chapter 1** (See **Fig 1.4**, on how the implementation of fallow practices may be related to specific descriptors which could be extracted by combining these two sources of data, hence helping with the conception of a mapping strategy based on such information.

3.1 Spatial analysis of reference data set

A first important step towards the proposed analysis is to assess the appropriateness of the reference data set to be used jointly with the available Sentinel-2 image time series, especially from the point of view of its spatial characteristics in the perspective of coupling such data with the specific resolution and level of details of the imagery employed.

For our case, the reference data set is built upon field data which has been collected over multiple years, suitably processed and stored in vector GIS format. We can find multiple overlapping entities (polygons) that cover the same area, but registered on independent field missions conducted on different years. As a consequence, inter-annual spatial discrepancies may emerge when considering data year by year (g̃rowing season) (see **Fig. 3.1**). Some of these spatial discrepancies could be caused by shifts in the acquisition protocol followed for generating the data set (see **Chapter 2**): for example, unaligned polygons could be associated to a change of boundary for a given field, whereas missing fields (i.e. "holes" in the multi-year sequence of records) can be explained by differences in the itinerary during the field mission or inability to recognize the specific land use (e.g. passage after harvest).

In general, we can identify a significant number of instances where previously registered fields are delimited differently in subsequent field missions, suggesting a farmer-driven reshaping of fields. Nonetheless, the overall surface of the Koumbia dataset which overlaps through multiple years might be sufficient for a reliable characterization of multi-year dynamics. However, in order to better assess this potential, and possibly even exploit field spatial dynamics in the conceptualization of a fallow mapping strategy, we conducted an in-depth analysis to quantify the proportion of fields that are exposed to these inter-annual reshaping, which will be presented in the next subsection.

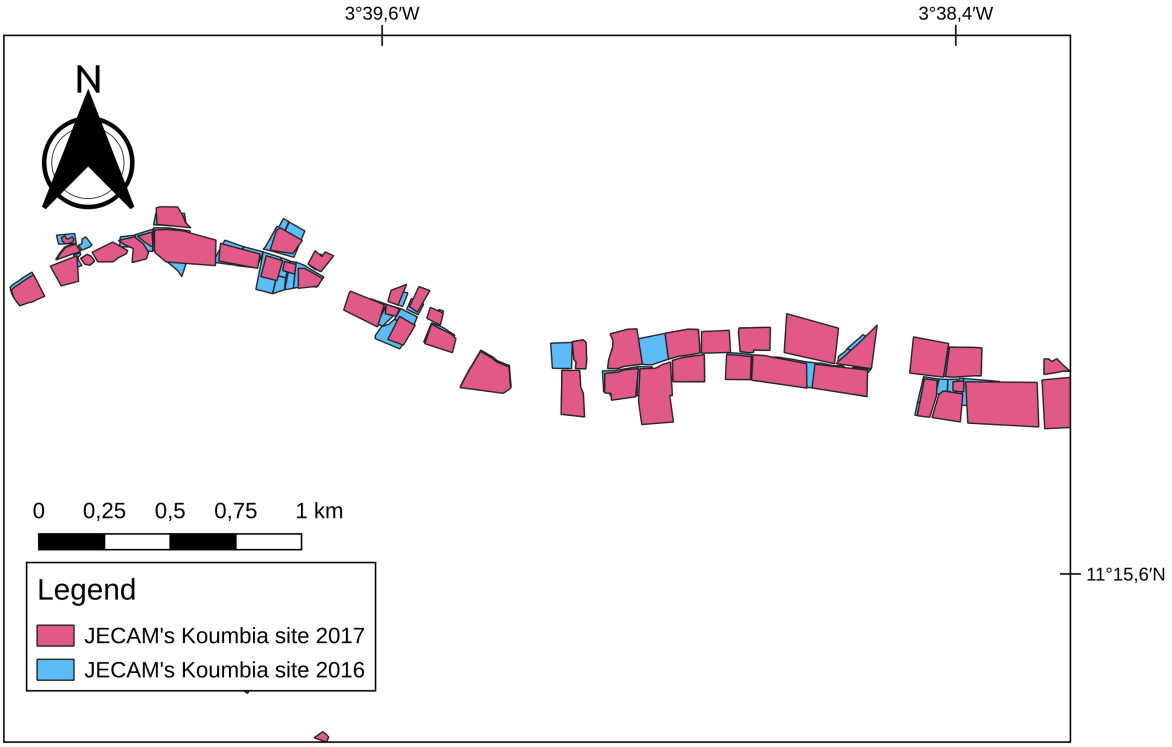


Figure 3.1: Example of two consecutive years of JECAM vectorial data with significant miss-matched surfaces.

3.1.1 Inter-annual spatial variation quantification

To quantify the proportion of fields that are affected by significant inter-annual spatial variations we adapted an analysis based on the comparison of segmentation maps over the area using the so-called Hoover metrics (see Hoover et al. (1996)). Just as a reminder, image segmentation is a process that takes an image as input (generally a multi-band image) and provides a partitioning of such image into disjoint objects (i.e. aggregates of pixels) covering the whole image extent. In other words, each image's pixel ends up belonging to one and only one object, and all pixels are assigned to a corresponding object. Many approaches exist to accomplish this task, some more appropriate than others for a use in remote sensing such as the "superpixel" approach proposed by the SLIC algorithm

(Achanta et al., 2012), statistical image processing methods (Gaetano et al., 2007), or region/object growing techniques (Baatz et al. (2000).

Hoover analysis allows the user to compare the objects defined by two different segmentation (in general, a *reference* one and a *machine-generated* one, but the comparison apply to any possible couple of segmentations) and identify if the objects of the two maps lying in a same area are *matching* (i.e. they cover the same area within a given tolerance), *over-* or *under-segmented* (i.e. a group of objects in one map lies in the extent of an object in the other one). In our case, we decide to somehow divert the use of this tool to compare two sets of objects issued by the rasterization of the Koumbia dataset over a given image lattice as a spatial reference. This allows for the detection of the temporal dynamics of fields (i.e. fragmentation through *over-segmentation*), and merging through *under-segmentation*) at least on the annotated sample to provide a statistical insight. The Hoover metric calculations operates as follows:

- two objects from the first and the second segmentation are *matching* (labelled as *correct*) if their intersection is above a given percentage (threshold *th*) of their respective area;
- a *over-segmentation* is detected if an object of the first map posses at least *th*-percent of its surface, distributed into a set of objects of the second map, whose total surface also must overlap the first object of at least the same percentage;
- a *under-segmentation* is the exact opposite of a *over-segmentation* : the overlap between a set of objects of the first map and a single one of the second is reciprocally at least *th*-percent of total respective areas;
- all other situations are simply marked as a *non-correspondence* (labelled as *missed*) between the corresponding objects.

Prior to hoover analysis, a series of pre-processing steps on the original spatial data sets have been necessary in order to provide a consistent comparison. In particular, after filtering out all fields not belonging to the *cropland* or *fallow* classes, for each year's data set:

- a spatial processing is performed to collate spatially all adjacent fields with the same crop type, to cope with cases in which a single field has been digitalized in two parts (probably as a legacy of previous digitalizations in the same area). First, fields are dilated using a small buffer size (1 m), then all overlapping polygons belonging to the same crop type class are *dissolved* (merged). A final negative 7 m buffer reduces fields to their final "core" surface;
- the resulting pre-processed data set is rasterized using the crop type or fallow class numerical identifier, over an artificial 5 m resolution grid (i.e., using as minimum sampling units a surface of 25 m^2), as a compromise between the preservation of the

original field shapes, the computational burden and the proximity with the Sentinel-2 spatial resolution;

- the connected components of the resulting raster (i.e., sets of adjacent pixels with the same class label) are independently labelled, in order to identify the different portions of field which are issued by a spatial split either forward or backward in time.

Finally, for each couple of subsequent year under analysis, a mask is applied to only include the common annotated surface. This is necessary to eliminate any residual bias due to a different application of the sampling/digitalization protocol from one year to another, which may lead to a considerable false detection of field boundary reshaping.

For the computation of the Hoover metrics, we make use of the tool implemented in the *Orfeo Toolbox* (OTB, v7.2.0). We then run the Hoover analysis over all couples of consecutive years of rasterized data sets, setting the overlapping threshold th by trial and error to 75%, which turns out to be a sufficiently restrictive value for detecting only significant changes in landscape. Note also that this technique provides a labelling of each independent object that can be consistently interpreted in both the senses of analysis (i.e., an over-segmentation in one sense is detected as under-segmentation in the other, etc.), thus making a causal analysis (through consecutive years in a single temporal direction) exhaustive w.r.t. all possible field boundary variation scenarios.

An example depicting the whole procedure is reported in Fig. 3.2: some typical fragmentation examples are in the left part of the sample area, with sets of monocrop *maize* fields in 2016 that are first remapped to single fields after pre-processing (see second row), in order to enable the detection of the field fragmentation happening in 2017 on the same surfaces (e.g., the big 2016 maize field in the top left part which is split in alternate *maize* and *cotton* fields in 2017). Also a merging is reported (bottom right of the area, in blue), on a small 2016 *maize* field which is merged to a big *cotton* field in 2017. The *missed* field is here considered too small (outside the 75% threshold) to be significant, while all green areas are "compatible" fields within the 75% threshold, net of any gaps in the digitalization protocol.

A statistical synthesis of the result, reported in **Fig. 3.2**, shows a ratio of inter-annual spatial variation, which is the sum of both types of possible segmentation discrepancies (i.e. over/under segmentation) reaching around 0.25 (or 25% of annual cropland in pixels) for five periods out of six, with significant lower values only for the 2013-2014 comparison, mainly due to a lack of common annotated surfaces. These results indicate a consistent active process of reshaping land surface use. Notice that periods related to year 2019 are not included since no vector data was initially available for this year. Note also that the given percentages only concern major modifications in field shape dynamics, due to the tolerance that we use in detecting coherent object splits and merges.

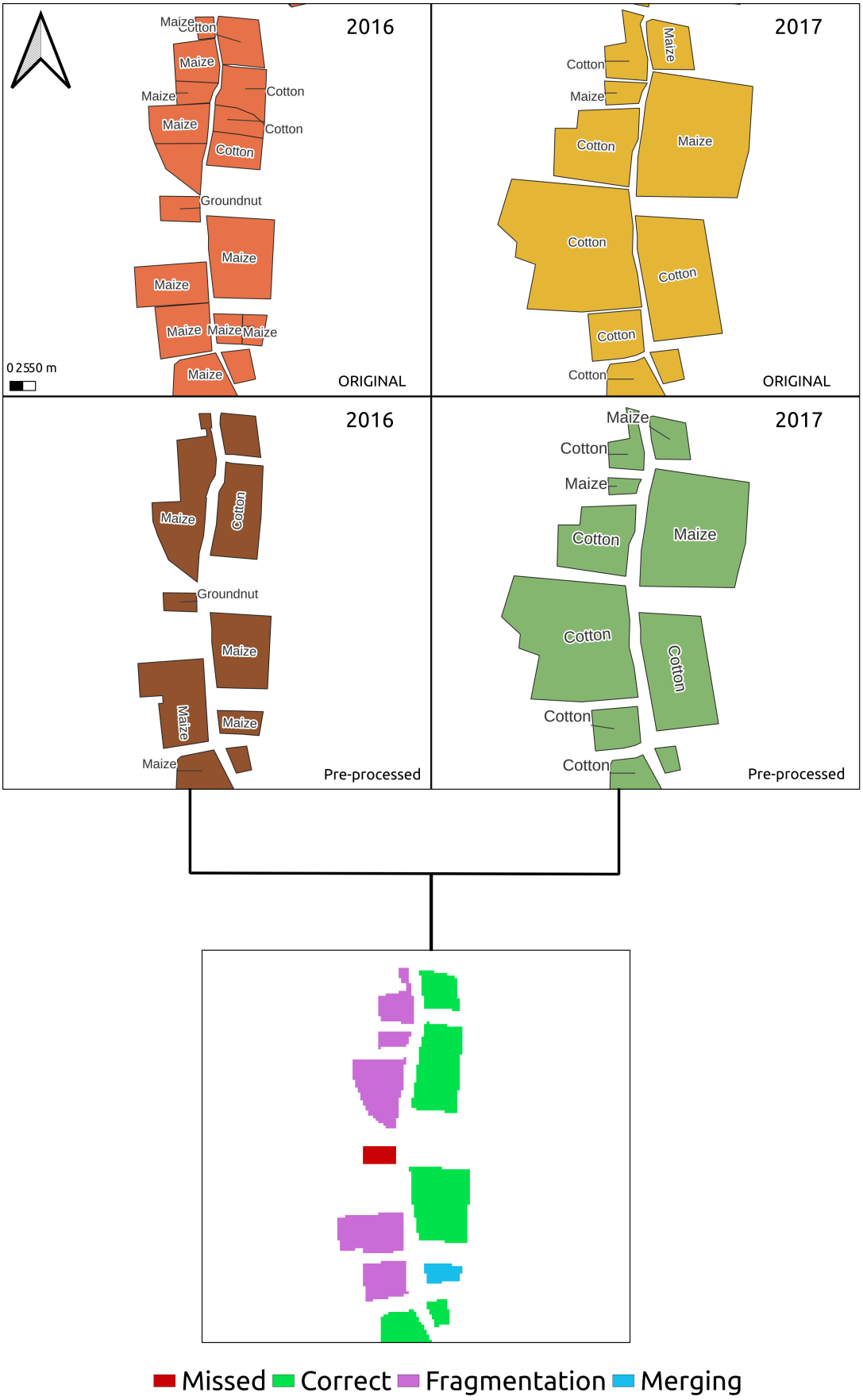


Figure 3.2: Example of inter-annual spatial variation results adapting Hoover segmentation comparison tool. (Top) Source JECAM fields, (mid) geo-processed JECAM fields prior to 5-meter rasterization and (bottom) color coded Hoover results using year 2016 as "ground truth".

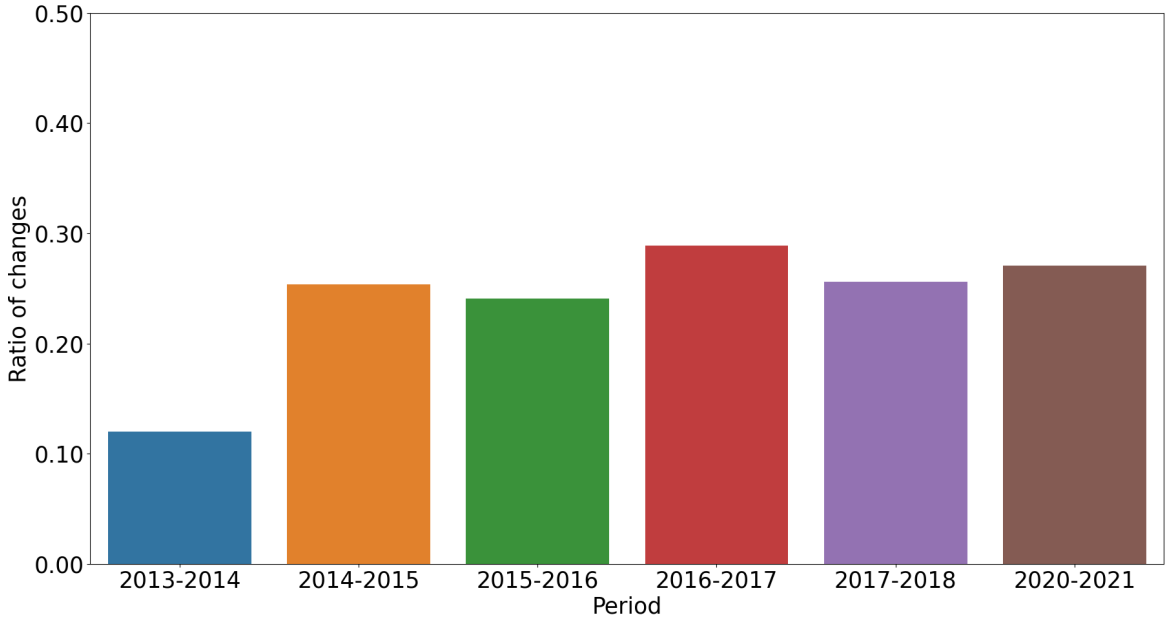


Figure 3.3: Ratio of surface subjected to inter-annual spatial variation (fields either splitted or unified). Fields are considered exposed to inter-annual spatial variation if at least 75% of its surface is affected. Values represent the total portion of rasterized cropland fields.

3.1.2 Reference data set spatio-temporal harmonization

A second main objective of this part of the work is to use the multi-year Sentinel-2 time series introduced in **Chapter 2** to characterize the effect of agricultural land use dynamics in the productivity of vegetated cover, particularly using the NDVI radiometric indices to identify trends related to the occurrence or absence of fallowing periods. To this aim, having a spatialized multi-year data set is a great opportunity, since this allows the aggregation of radiometric indices at the plot scale, enabling a more robust characterization through zonal statistics. This aspect is fundamental in such agrosystems, due to the strong intra-field variability which would make a point sampling strategy too uncertain.

Nonetheless, the analysis presented in the previous section raises an important issue in the perspective of performing a robust characterization of fallow surfaces using the multi-year satellite image time series. In fact, the reference data utilized lacks rigorous spatio-temporal continuity. This inconsistency arises not only from missing fields reported for certain years but also from the local reshaping of agricultural boundaries. For this last reason, when using the original polygons from the Koumbia data set, fields that are registered as *fallow* (or other LULC classes for that matter) in a given year, although significantly overlapping with their homologous, are with high probability reshaped into a new field disposition in another year, if not omitted due to a missed visit during the field mission. If the latter problem eventually causes a simple missing value in the series of recorded land uses, the former on its side generates ambiguity about the surfaces that must be used to aggregate zonal statistics from the satellite images. For these reasons, we decided to construct a spatially consistent vector data set allows the multi-year analysis

of the agricultural land use but at the same time preserves most of the original JECAM data.

The idea is to enforce local consistency by identifying portion of land to which we can assign one, and only one, *land use trajectory* across the different years of the data set, possibly with holes due to missing data, but with a non-zero surface. Since the missing field situation is quite recurrent in the data set (without mentioning the completely missing information on fallow land in 2019, see **Sec. 2.2.2**), accepting partial sequences is necessary as it would maximize the probability of gathering a large enough sample to compute statistics on crop sequences and crop/fallow cycles.

To build such a set of *consistent land units* we performed a specific geo-processing workflow in vector GIS format, based on the subsequent intersection of field polygons through the different years of the data set.The process is iterative, based on the subsequent cumulative intersection of each year's set of fields (*layer*) where, at each iteration, we make two different choices according to the result of the intersection: (i) for overlapping fields, we only keep the intersected portion(s) in order to have coherent surfaces for the multi-year Sentinel-2 based characterization, while (ii) non-overlapping (disjoint) fields in both the intersected layers are simply copied to the output layer, under the assumption that their characterization is still possible at the price of a hole in the sequence of land uses. A scheme of the procedure is depicted in **Fig. 3.4**:

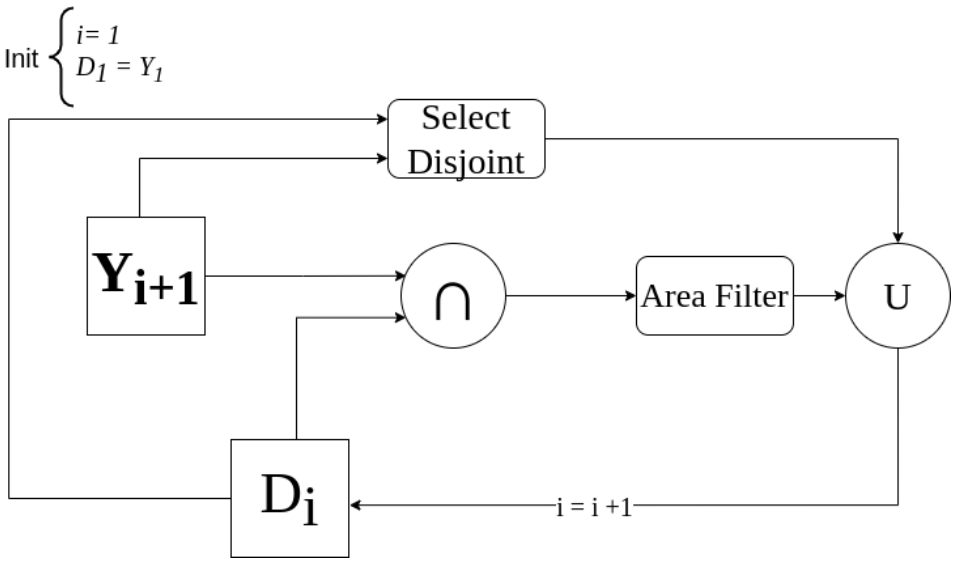


Figure 3.4: Scheme of the spatio-temporal harmonization procedure conducted over JECAM dataset

given a sequence of N available years $\{Y_1, \ldots, Y_N\}$, at iteration $i = 1$, the output D_1 is initialized with fields from the first annual layer Y_1. At each iteration, the current result D_i is intersected with the next layer in the sequence, Y_{i+1} and a filter based on an area threshold (set to $200m^2$) is applied to remove small residual fields that cannot be reliably characterized at Sentinel-2 resolution ($100m^2$ per pixel). In parallel, the two layers are compared to recover disjoint fields and add them to the next result D_{i+1}. Procedure stops

at iteration $N-1$, then a morphological *opening* is performed on the final output layer $D = D_{N-1}$ by applying a negative buffer of $5m$ followed by a positive buffer of $3m$. This last step jointly provide a filtering of small residual areas and a remodelling of the resulting fields to more regular shapes by eliminating thin appendices which may introduce noise when computing zonal statistics from the Sentinel-2 time series but cannot be filtered with a simple area threshold.

Finally, this procedure yields a single set of non-overlapping polygons containing the underlying LCLU multi-year data. A zoom on a clip of the resulting data set is presented in **Fig. 3.5**, and compared to the original layers from two specific years.

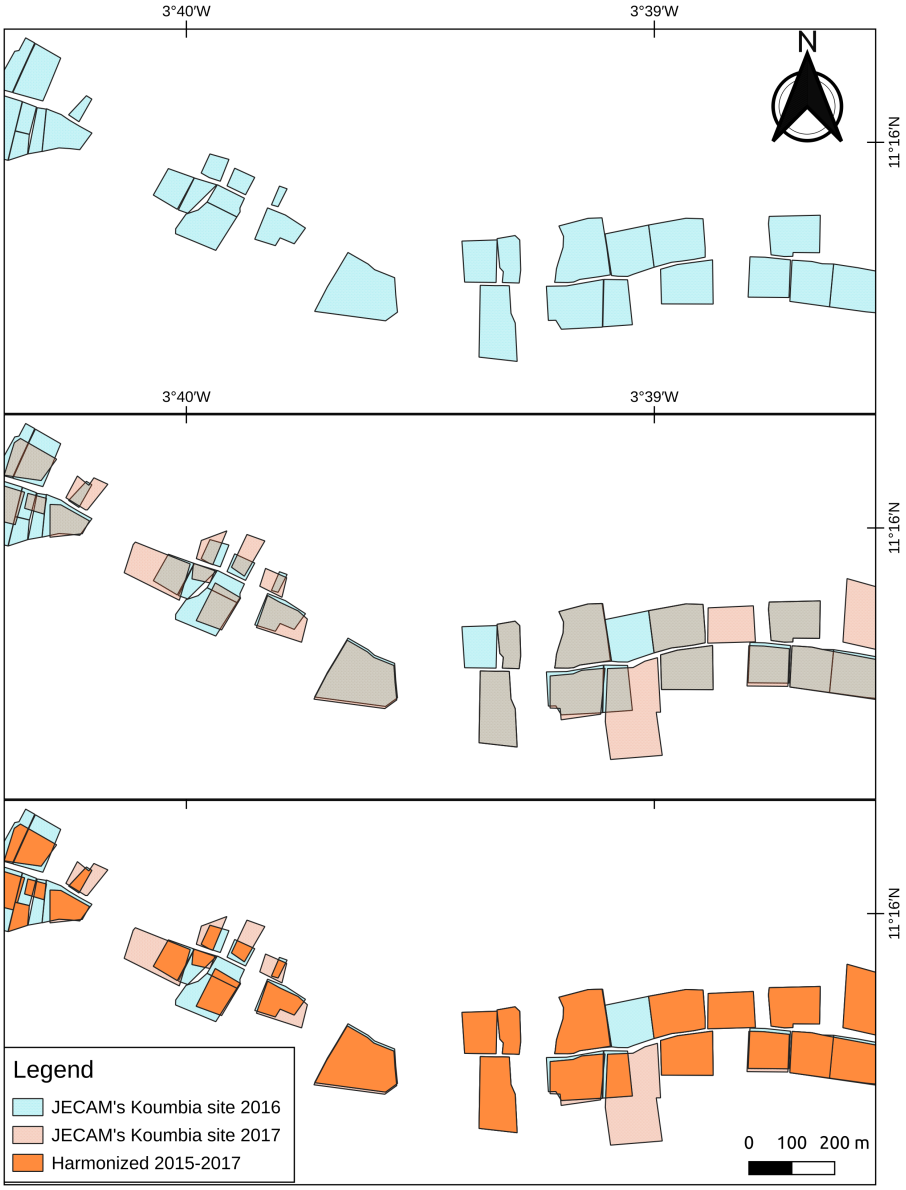


Figure 3.5: Example of resulting processed JECAM's "harmonized surfaces" compared to two consecutive years of JECAM vectorial data with significant miss-matched surfaces. Grey color indicates overlapping surface.

3.2 Analysis of multi-year fallow NDVI profiles

Now that we dispose of the coherent set of reference surfaces to attribute with remote sensing based indices, it is time to select more specifically the information we derive from our input multi-year satellite image time series. In order to inspect the vegetation dynamics related to the different uses of the cropland in our study site, we employ the Normalized Difference Vegetation Index (NDVI, Rouse et al. (1974)) as a base spectral feature that could allow to identify possible discriminating insights, which is calculated as follows:

$$NDVI = \frac{NIR - RED}{NIR + RED} \tag{3.1}$$

Where NIR and RED represent near-infrared (wavelength $\tilde{7}$60-1500 nm) and red (wavelength $\tilde{6}$00-750 nm) bands respectively. In our case we use Sentinel-2 bands *B4* (central wavelength 665 nm) and *B8* (central wavelength 842 nm) for *RED* and *NIR*. NDVI is typically used as an indicator of the vegetative activity (or level of "greenness"). As any *normalized difference*, values of the NDVI range from -1 to +1. Values closer to +1 indicate high levels of presence of green, healthy vegetation, whereas values close to 0 are usually associated with the absence of any active vegetation. Negative values are more typical of humid surfaces and water bodies. This index relies on the fact that photosynthetically active biomass reflects a high portion of the incoming sun radiation in the *NIR* region whereas the opposite is true for the *RED* region. It is useful to analyze NDVI values along the cropping season, as this index varies along the season with high positive values during peak growing season where most of vegetative development occurs, and low values close to zero when there is none or low vegetative growth. Hence, most of vegetative differences described by this index cannot be captured with a single measure, but rather through its dynamics along the cropping season. Thus, we use NDVI profiles (a NDVI for each image of the time series) as a primary tool to analyze the dynamics of this indicator for all JECAM's LULC classes that might present a vegetative cover (i.e. cropland, forest, etc) and dismissing less relevant classes such as water bodies or urban surfaces.

3.2.1 Class-wise NDVI profiles

A first inspection of NDVI profiles at an annual scale, as in the example shown in **Fig. 3.6** for year 2017, reveals very subtle differences in NDVI intra-seasonal dynamics among the different LCLU classes considered with the exception of *woody* class. The *woody* class presents an average NDVI more stable along the period considered, with higher values during the dry season. Fields that make use of fallow practice present similar NDVI profile dynamics as other major crop classes such as *cereal crops* or *leguminous crops*, with a peak by the end of the rainy season. This is to be expected in rainfed agricultural systems, since most of vegetative growth is likely to be conditioned by the availability of water. Hence,

both cropland and other types of natural vegetation spaces present a growing NDVI value once the rainy season begins. In the example shown in **Fig. 3.6** peak NDVI is reached in October, however considering that most of the precipitations occur between May and September we can assume that the likelihood of cloud-free images during this period is rather low. In fact, despite using the gap-filling strategy to mitigate this issue, we can notice that the NDVI for all classes in the period between July and October follows a rather linear trajectory, most likely as a consequence of this image processing technique itself. In an ideal case, where all images present no imperfections (i.e. no clouds, no shadows, no noise, etc), we could expect a more "bell-shaped" NDVI curve with a less pronounced peak. Thus, the NDVI dynamics captured are strongly conditioned by cloud presence for our study site, which ends up rendering any image classification task more difficult.

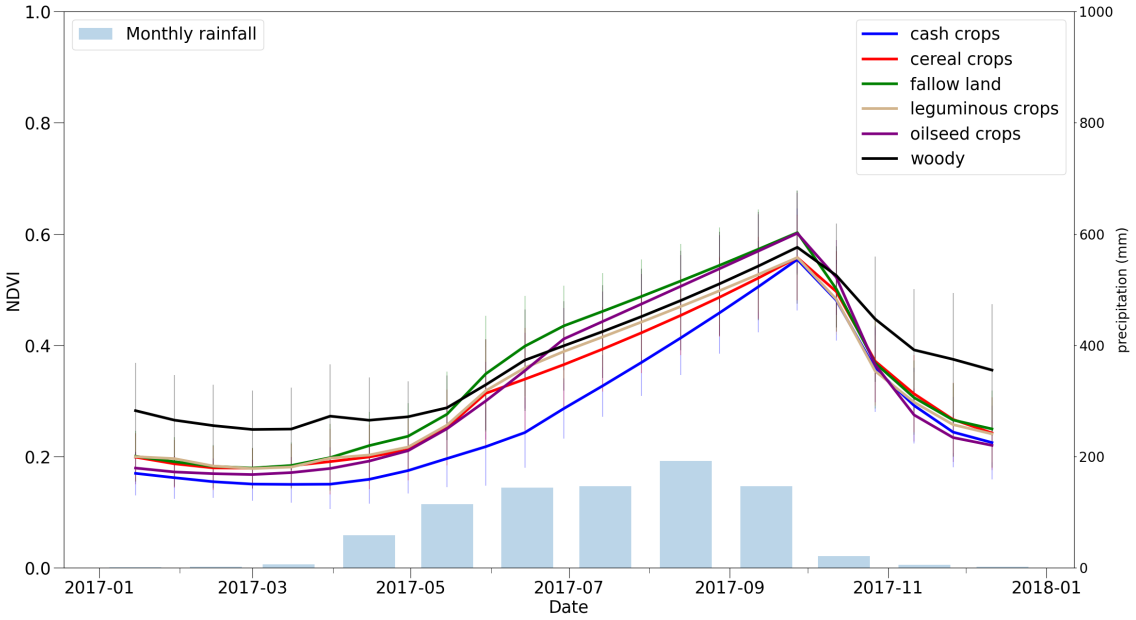


Figure 3.6: Average seasonal NDVI profiles calculated with gap-filled Sentinel-2 data and JECAM's Koumbia subset. Rainfall data obtained from CHIRPS monthly data (Funk et al., 2014). Vertical lines represent standard deviation for each LCLU class. Adapted from (Castro Alvarado, 2022).

Consequently, annual mapping strategies for differentiating any type of vegetative class under this rather unfavorable radiometric context are much likely bound to fail, with a possible exception for the woody class. This class is composed of a less senescent type of vegetation, which translate into a characteristic higher NDVI values during dry season compared to rest of vegetative classes, and where cloud presence is reduced.

As for the case of *fallow* class, there are no apparent contrasting NDVI features during growing season. Nonetheless, the effects of such agricultural practice are likely not detectable during the first fallowing season, but rather during the subsequent cropping seasons after one or more years of fallow (or "post-fallow seasons"). Hence, according to the known properties of soil regeneration associated with the fallow practices, one could

reasonably make the hypothesis that regenerative effect of fallow practice are more likely to be reflected in higher NDVI values through the course of multiple consecutive years of non cultivating a given land. Thus, by making use of the reference *consistent land units* whose generation process is described in **Sec. 3.1.2**, we generate average yearly NDVI profiles for *fallow* samples, but this time differentiating fields according to the duration of the fallowing process (see **Fig. 3.7**). According to this figure, NDVI results suggest that, with the exception of five-years old fallow land, more than one consecutive year of *fallow* present a higher NDVI maximum peak, reached between day 240 to 300 of the year (i.e. between late August and late October). We can see that four-years old *fallow* land has the highest NDVI peak and that there is not clear distinction between two and three-year-old fallow land. A slight difference between the moment when the peak of NDVI is reached is appreciated between two and three-years old *fallow* land. However, it has to be noted that the duration-based *fallow* typology considered in **Fig. 3.7** is extracted from JECAM's data coming from different years (e.g., a three year old fallow can be between 2015-2017 or between 2016-2018). Furthermore, the longer the *fallow* period, the less data are available for that given *fallow* typology. Thus, *fallow* phenology is strongly dependent on the amount of available data and on the climatic conditions under which this data has been acquired. Consequently, special precaution is needed regarding the interpretation of these results given the low statistical relevance due to the reduced amount of fields with a confirmed duration of at least two consecutive years. Another extra factor that is not being accounted in these results is the effect of location, or more specifically soil conditions which can further add intra-class variability.

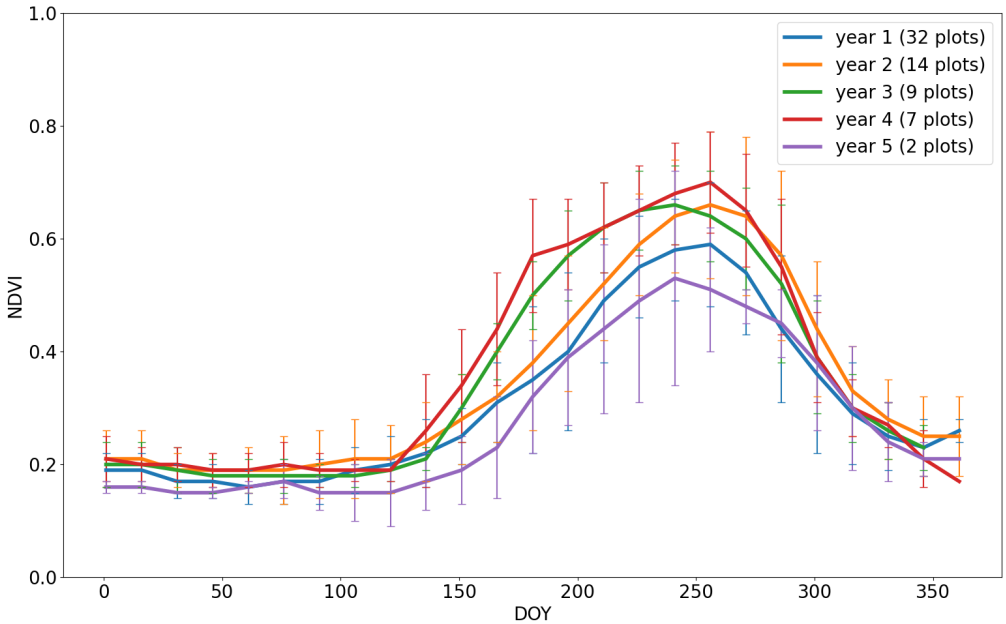


Figure 3.7: Average seasonal fallow NDVI profiles calculated with gap-filled Sentinel-2 data and processed "reliable surfaces" (see **Sec. 3.1.2**)

3.2.2 Field level multi-year NDVI profiles

In an effort to further analyze *fallow* practice using NDVI as a describing feature and reduce the effect of weather conditions, we also conducted a multi-year NDVI analysis at the field level. Again, we use the reference consolidated surfaces presented in **Sec. 3.1.2** as spatial reference for generating per field level multi-year NDVI profiles (see **Fig. 3.8** and **3.9**).

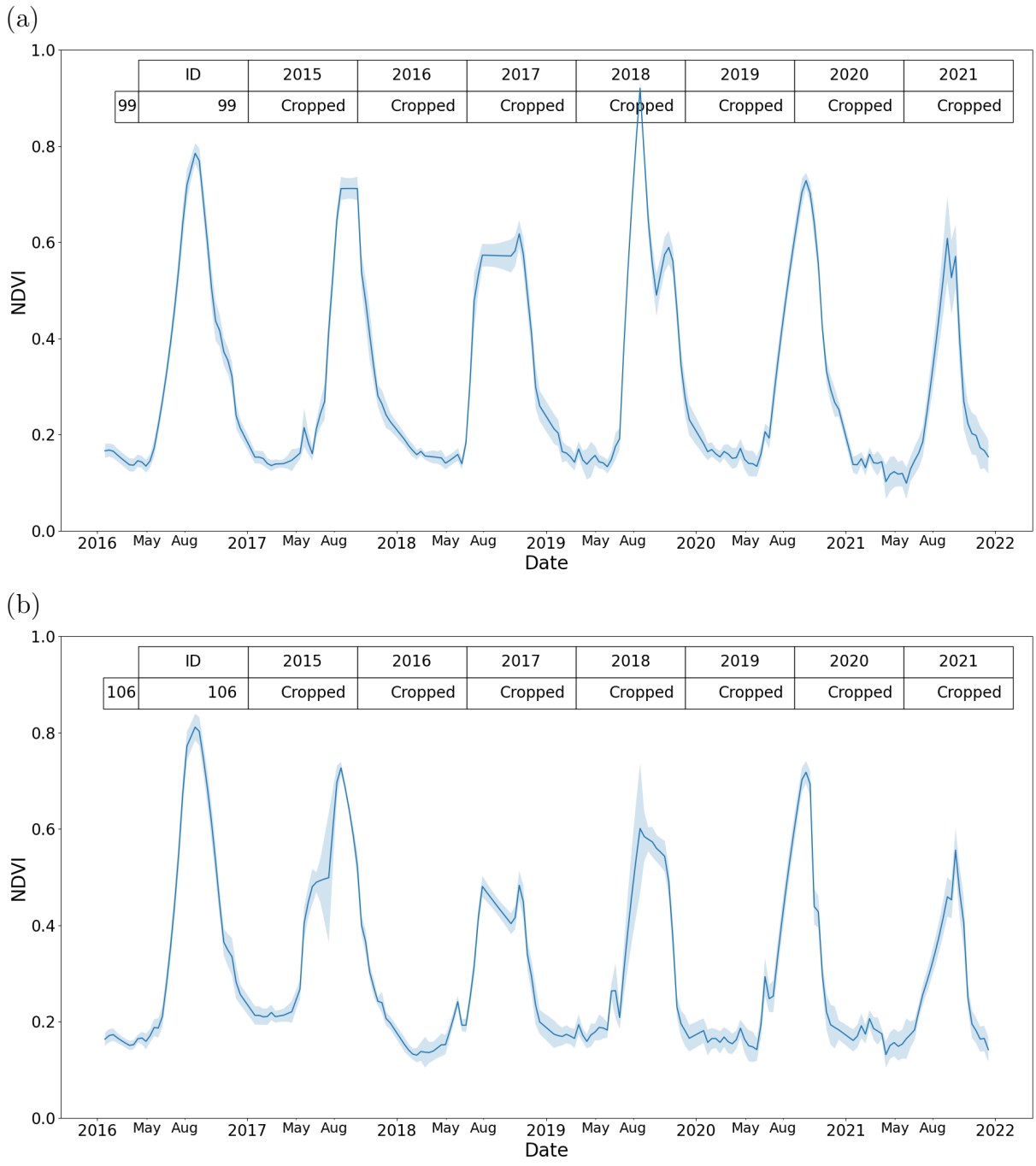


Figure 3.8: Average multi-year NDVI profiles at field level extracted over processed JECAM's spatio-temporally consolidated surfaces (see **Sec. 3.1.2**): examples for monocrop fields. Light blue shaded area indicates the standard deviation.

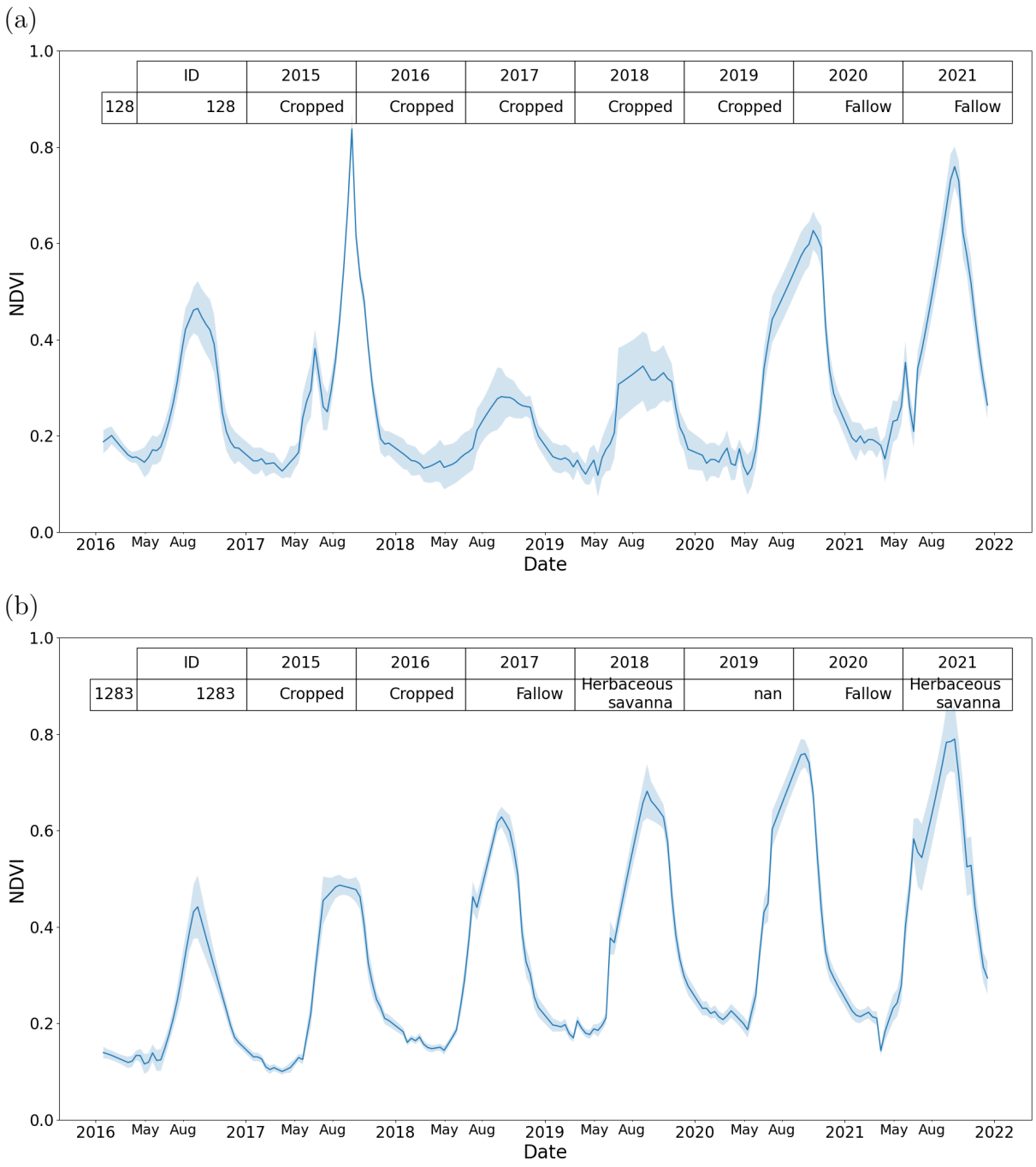


Figure 3.9: Average multi-year NDVI profiles at field level calculated over processed JECAM's "spatio-temporal stable fields" (see **Sec. 3.1.2**) examples of fields including *fallow* practice. Light blue shaded area indicates the standard deviation.

In **Fig. 3.8** we can see two different fields that have been cropped every growing season for the period 2015-2021. If we consider year 2016 as a reference season for comparing NDVI dynamics, where both fields were cropped with maize and have similar NDVI behaviour, their multi-year profile also follows a similar behaviour for the rest of the period of analysis. Noticeably, year 2018 and 2021 have significantly lower NDVI peaks when compared to 2016 season. The differences between these two fields during that two given years are likely the result of differences in the chosen crop grown those years, as

well as the specific soil properties that differentiate these two fields (if any). For instance, field number *99* was cropped with maize and cotton for years 2018 and 2021 respectively, whereas field number *106* was cropped with soybean in both years. In this way, although both field could be integrated into a single "cropped" class, we identify strong NDVI differences depending on a series of pedo-agroclimatic circumstances. However, most importantly, we appreciate a global trend of loss in vegetation productivity over the years, especially if we keep annual rainfall conditions into consideration.

Conversely, if we analyze instances where cropped fields have been alternated with *fallow* periods (see **Fig. 3.9**), we observe what appears to be a cumulative positive effect on maximum NDVI values during and after cropping activity has been paused. Moreover, we detect instances where a given land known to have been cropped in past has been registered as a type of natural vegetation space (*Herbaceous savanna* in example), highlighting the difficulties of differentiating non-active agricultural land as there are no clear markers that allow identifying land beyond all doubt as such (here, possibly an annotation error due to the absence of knowledge on past activity from the field operator). When comparing NDVI results from both figures (**Fig. 3.8** and **3.9**) on year 2021 we observe a significant difference between fields that have been cropped and *fallow* land, where cropped fields have a NDVI peak close to 0.6 and *fallow* land is closer to 0.8 for the same year. We also value a modest early development of vegetation in *fallow* land, observing NDVI increases as early as May, whereas cropped fields had a shorter cycle which started in late July-August. This suggest a possible delay in the sowing dates of such actively managed fields, with respect to the spontaneous development of vegetation over *fallow* land, the former basically not taking advantage of early precipitations at the beginning of the rainy season.

The selected examples depict some of the different trajectories a given land can encounter along the period considered, where in some instances there are significant changes in seasonal peak that coincide with periods of changes in the type of management done during the season which, in this case, are either cropped or fallow. Nonetheless, the illustrated trends/differences between land use types along multiple consecutive years are a rather uncommon behaviour among all the fields analyzed. This can be explained by the highly heterogeneous landscape which is the result of the presence of weeds and heterogeneous cropping practices (e.g. seeding and harvesting are done during multiple weeks) which in turn produces a rather smooth NDVI signal.

3.3 Agrosystem related drivers for fallow practice

As we discussed in **Chapter 1**, there are multiple factors that can imply the emergence of *fallow* practices, some of which can be related to the loss of vegetation productivity (or similarly, the will to restore fertility), and characterized through the analysis of NDVI dynamics like the examples we provided in the previous section. In this section we will analyze other possible drivers, more related to external agro-systemic factors than to the

cropping conditions, which may favour the emergence of *fallow* practices and help define *a priori* knowledge about their relative location in the landscape.

3.3.1 Fallow practice and soil types

One of the first characteristics to test is the relationship of *fallow* land with regards to soil properties. Taking into consideration that most of cropland in our study site (and most of Sub-Saharan Africa for that matter) is built upon rainfed agriculture, one of the main limiting factors for crop growth is likely to be water availability. Available water for plants is dependent on both precipitations and *soil properties*. For example, it is reasonable to assume that shallow soils have a lower soil water retention, i.e. water is not stored in soil in sufficient quantity and would tend to quickly evaporate into atmosphere, whereas in sandy soils water would tend to filter down to depths where plant roots cannot reach this water. From an agricultural point of view, these field might be considered less fertile and hence less interesting to work upon, hence more likely to be left uncultivated when conditions are not optimal (e.g. low precipitation, reduced amount fertilisers, etc). Consequently, we can make the hypothesis that specific soil types are linked to the emergence of fallow land.

In order to test this hypothesis, we make use of the JECAM data set including data from the period 2013-2021, where, thanks to the available annotations, the original classes have been grouped into four cropland major crop classes (*leguminous*, *oil seed*, *cereals* and *cash*), a *woody* class and a separated class for *fallow* land and analyze the soil types within each JECAM field. As for soil data, we tested both a freely available product which covers the area, namely the 30 meter ISDA USDA soil texture data for the top 20 cm (Miller et al., 2021), and a 10 meter *ad-hoc* soil type map that we generated making use of Sentinel-2 imagery during dry season (see*Dry season soil map generation procedure* in annex). Then, we calculated the proportion of each soil type present at field level (see **Fig. 3.10**).

As shown in **Fig. 3.10**, for the case of ISDA soil data, most of the cropland in the study case is dominated by a few soil type classes, which correspond to the USDA soil textural classes *Sandy Loam* and *Sandy Clay Loam*. For the case of our custom soil classification, obtained by clustering areas according to the visible soil properties during the dry season, what we have is basically a purely radiometric categorization, offering a clear evidence of contrast among different type of soils but with no pedological clues. We hence limit the categorization to a simple numeration of the detected soil types. As for the study area, three types of soils, namely *soil type 0*, *soil type 3* and *soil type 4*, are the most prevalent classes across JECAM cropland, with a reduced presence of *soil type 3*. Notice that, *soil type 2* is not present in any cropland group. Moreover, it appears that that *soil type 3* could be linked to woody species with a clear predominance in *woody* class, hence revealing that this particular class may more reasonably represent a more generic woody class rather than a soil class, since on these areas no soil is visible also during the

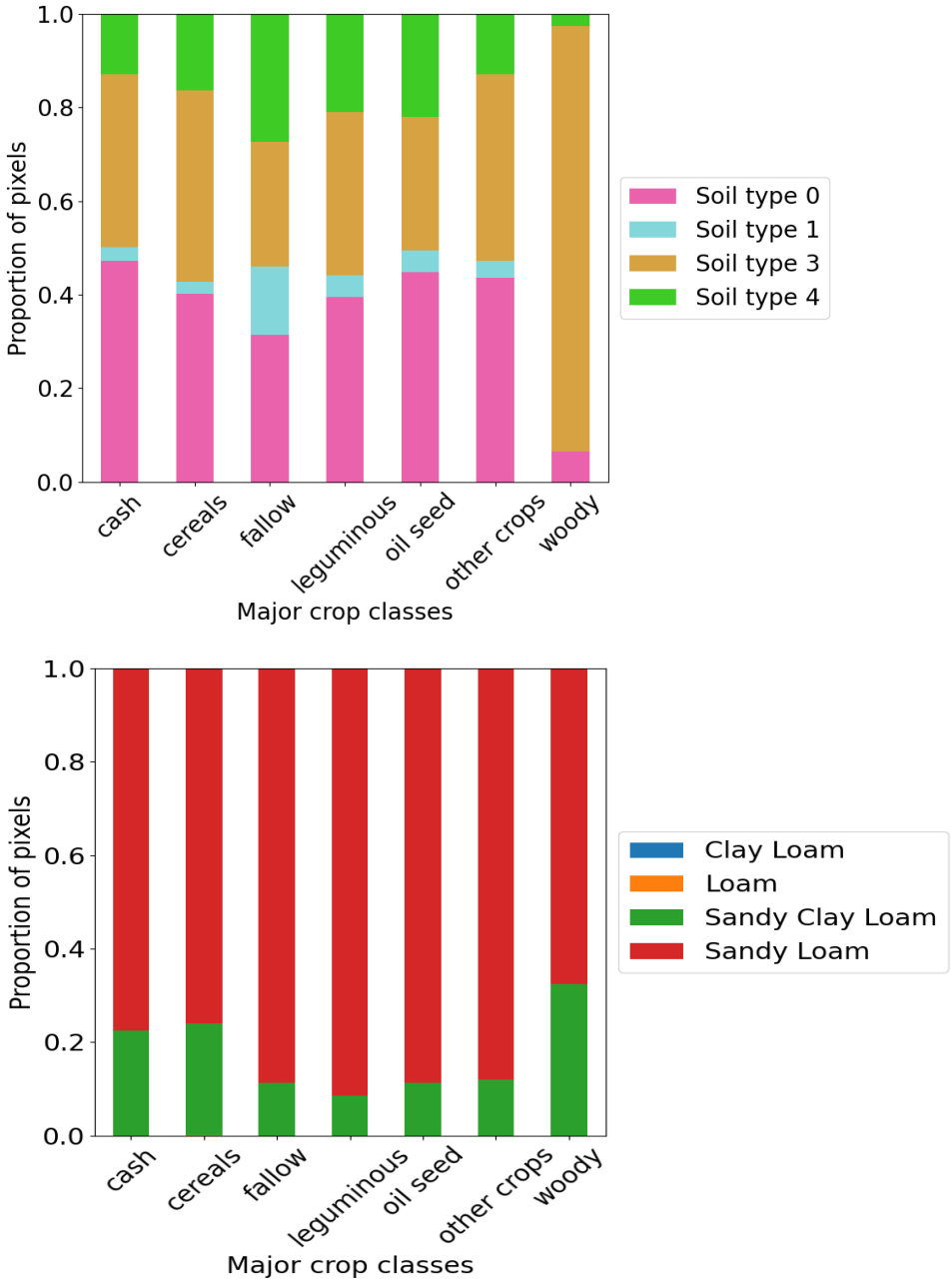


Figure 3.10: Proportion of soil types within JECAM fields data for the period 2013-2021, with respect to ISDA soil data (Miller et al., 2021) and *ad-hoc* soil typology (see **Sec. 7.5**).

dry season. The significant presence of *soil type 3* in the rest of cropland groups might be induced by a high presence of trees in crop fields.

No significant distinction can be observed for *fallow* land with respect to other types of LULC classes considered independently of soil classification product employed. Notice that for the case of ISDA soil map, there are two extra soil type classes which are present in the study site (i.e. *Loam* and *Clay Loam*), yet not represented in the reference ground truth data employed. There was no sufficient evidence to support the hypothesis about a possible correlation between the type of soil and the emergence of fallow, as tested in this section. However, it is unclear to which extent these results are influenced by the sampling

protocol followed during JECAM field missions. When considering the distribution of field included in source JECAM data, it is clear that most of sampled fields are limited to those close to main roads for accessibility purposes. In this manner, the representativeness of a potential link between soil type and fallow might be hampered by the spatial distribution of sampled fields in reference data. "For more details on the JECAM sampling strategy and implementation, please refer to Jolivot et al. (2021).

3.3.2 Fallow practice and its relative location

Another spatial dimension that we can analyze as a potential descriptor of *fallow* practice is the relative location of fallow land with respect to populated areas. One factor that describes smallholder agriculture is its rather pragmatic approach towards agricultural decisions. Thus, the amount of work dedicated to a given land will be determined by the expected return of a given field, favoring land known as, so to speak, "fertile". Nonetheless, these decision are strongly influenced by the availability or lacking of workforce and mechanized labor. In this manner, under a context where workforce is reduced, not renewed and progressively older as a consequence of young population rural exodus, it is likely that most of agricultural activity is clustered around the vicinity of villages. Hence, less attractive fields, from an agricultural point of view, would likely be located away from home field and might be left uncultivated more often than closest fields.

In order to explore this hypothesis we make use of the High Resolution Settlement Layer (HRSL) (Lab and for International Earth Science Information Network-CIESIN-Columbia University, 2016), containing geo-spatialized estimates of population density for Burkina Faso at a 30 meter resolution for year 2015. We assume that the value stored in each 30-meter pixel might have varied along since 2015, but the location of the population clusters (i.e. *settlements*) has been rather stable across time. Thus, we produce a proximity map which calculates the distance between target pixels (i.e. settlement pixels) and overlay unaltered JECAM data for the period 2013-2021 to calculate per field settlement distance statistics (see **Fig. 3.11**).

Results shown in **Fig. 3.11** show no apparent significant differences among cropland classes considered in JECAM dataset with the exception of *woody* class which includes a series of minority tree-based crops. Excluding *woody* class, all interquartile ranges (i.e. box length) overlap between classes, with mean and median distance for all classes reaching 9.03 Km and 7.45 Km respectively. Most of cropland is hence located below the 10 Km mark from the nearest human settlement as described in the HRSL, with a significant portion of cropland reaching a distance of more than 15 Km and an average distance to the nearest settlement of 6.3 Km. Results did not highlight higher distances for *fallow* land to human settlements, hence rejecting again the proposed hypothesis. Nonetheless, there are multiple instances of cropland fields located far beyond the 15 Km mark, suggesting that agricultural activity might extend beyond the area sampled in JECAM dataset. As mentioned in previous section (see **Sec. 3.3.1**), spatial representativeness might be ham-

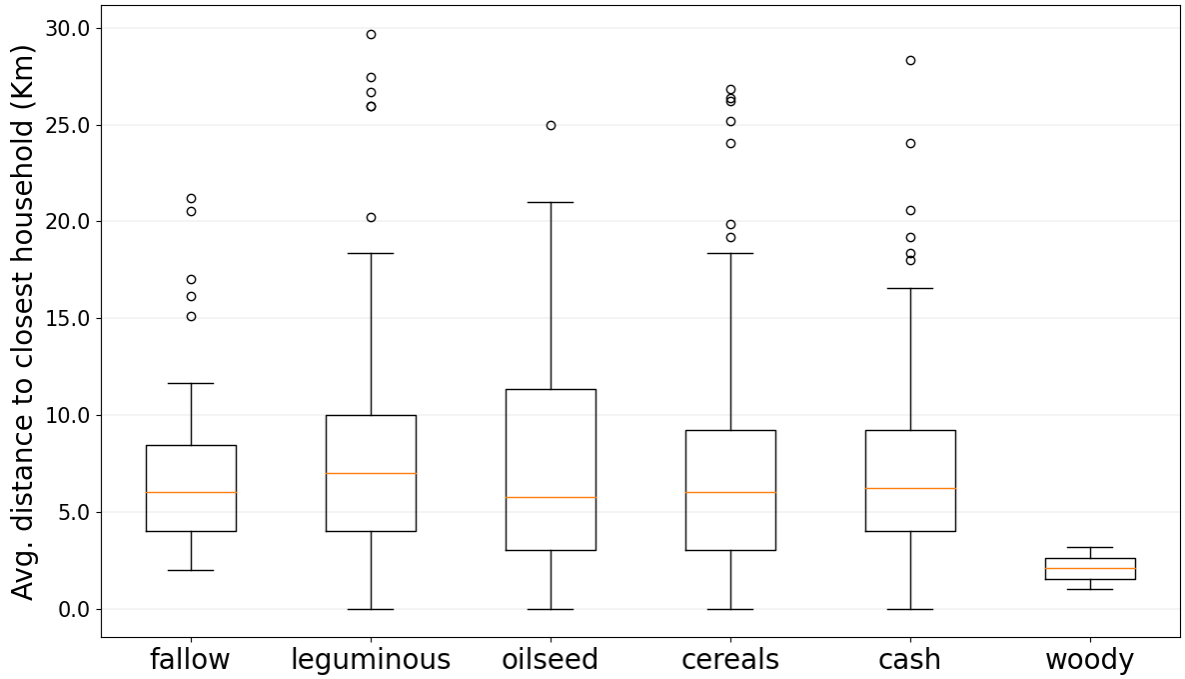


Figure 3.11: JECAM cropland fields distance to human settlements derived from Lab and for International Earth Science Information Network-CIESIN-Columbia University (2016)

pered by JECAM sampling protocol, where most remote fields are not sampled due to accessibly issues and potentially not including less active cropland. As a consequence, minority classes such as *fallow* land class could be negatively affected by a reduced coverage during the field missions. Not to mention that main classes are only sampled alongside main roads hence potentially missing a significant portion of cropland that are closer to villages which can translate into a distortion of resulting distances.

3.4 Conclusions

On this chapter we explored different aspects for describing cropland in our study site making use of JECAM geo-spatialized data. From an spectral perspective, there are no clear differences between main crop classes when analyzing annual NDVI profiles. Multi-year NDVI profiles calculated at field level suggest a cumulative effect of *fallow* land when conducted during multiple consecutive years. However, due to a highly heterogeneous spectral context, rule-based modelling does not appear viable under the agroclimatic conditions considered. When analyzing the spatial distribution of JECAM data set, it was detected a significant inter-annual spatial variation, reaching up to 15% of JECAM fields affected by process of either merging or fragmentation, suggesting active agricultural field reshaping (i.e. new field boundary definitions). JECAM's sampled fields gathers around main roads, hampering the representativeness of the field distribution. Consequently, hypothesis related to the relative location of *fallow* fields cannot be verified, further justifying our major investment toward a more data-driven approach since no explicit

modelling seems possible on the study area.

Chapter 4

Annual mapping strategies

In this chapter we analyze the potential of automatic image classification methods to discriminate fallow land from other types of land uses on a yearly basis using seasonal information. We implement two different strategies for discriminating *fallow* land : (a) a more classical supervised classification workflow, and (b) the unsupervised methodology proposed in Tong et al. (2020) originally developed for mapping *fallow* land across the Sahel band.

4.1 Direct Detection of Fallow Class

One of the first strategies for detecting non-active agricultural land is to attempt to detect it as another land use/land cover class comparable to other types of uses such as specific crop types (e.g. maize, cotton, etc) or other non-cropland classes (e.g. forest, savannas, built-up spaces, etc). This approach is worth testing in our case due to the presence of samples in the JECAM data set explicitly annotated as fallow that we can use to assess it. In principle, such approach could yield accurate results as long as (1) the fallow practice is associated with highly contrasted spectro-temporal features (commonly making use of SITS) and, (2) one disposes of sufficient reliable reference data for each class to verify predicted results. Hence, when there is a clear and stable delimitation of the agricultural land, this tasks is reduced to identify those areas within cropland that present clear signs of the absence of human activity. Thus, a clear understanding of local *fallow* practices and a cropland delimitation (usually achieved through the use of cropland masks products) are necessary to better isolate the practice in time and space. The methodological approach implemented will depend on the general agro-climatic context and on the characteristics defining the target *fallow* practice. For instance, if fallow land is known to be conducted in conjunction with weeding activities, one could expect a highly contrasting field with low vegetation growth compared to the rest of cropland. In this case a rule-based system might suffice to differentiate this practice, whereas in regions where this practice is not clearly defined or it does not present highly distinguishing features automatic classification systems relying in machine learning algorithms might be more adequate.

As mentioned at the introduction of this chapter, we will analyze two different approaches, a more "legacy" supervised strategy where fallow land is one extra cropland class and where we make no use of any cropland mask, and another unsupervised approach that replicates fallow mapping procedure proposed by Tong et al. (2020) for which in principle no explicit ground truth on fallow practices is needed. Both of these strategies are described below and will establish the baseline of to which compare other strategies tested in posterior chapters.

4.1.1 Supervised annual fallow mapping approach

In this strategy we will perform image classification based on a 6-class nomenclature: *cropped*, *fallow*, *water body*, *evergreen*, *herbaceous savanna* and *non-vegetated surfaces*. *Non-vegetated surfaces* is composed of JECAM's *built-up* and *bare soil* classes, whereas *evergreen* is composed of *forest* and *savanna with trees* classes. Given a lack of cropland products with sufficient accuracy at local scale for our study site, we decided to, instead of using a third party cropland mask, include non-cropland classe into the classification process. The number of data points (or pixels) available once reference data is rasterized using Sentinel-2 images as spatial reference (10 meters) is shown in **Fig. 4.1**. It can be seen that the total number of data points varies significantly among the classes considered, yet the distribution of these into cross validation folds is rather balanced.

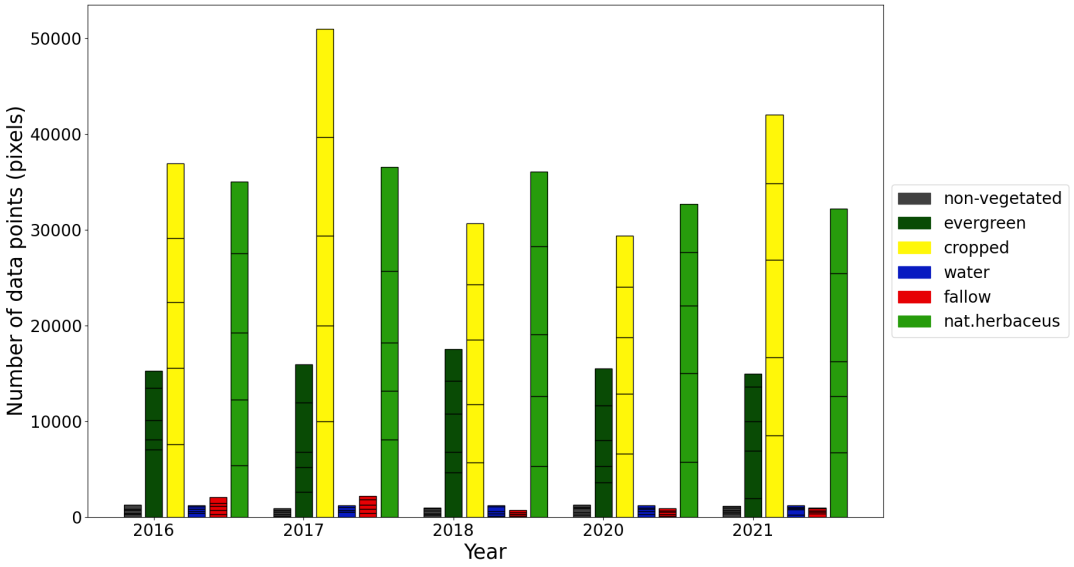


Figure 4.1: Number of data points in pixels (after rasterization using the Sentinel-2 10-m grid as spatial reference) for all years and separated by LCLU classes. Bar splits represent the proportion of pixels employed in each cross validation fold, see **Chapter 2**.

Model training process followed the standard procedure described in **Chapter 2** (see **Sec. 2.3.3**) and performed annually for the period 2016-2021, using both image classification algorithms Random Forest and TempCNN. Note that, in order to have comparable models, we use as input data Sentinel-2 reflectance bands and indices only, thus rainfall and other non-decametric features are not used here. This can be explained by the fact

that for TempCNN to be able to exploit temporal relationships between features, these require an equal time-step distribution for all variables. Thus, any extrapolated data could hamper model performance. In our case we consider that gap-filled Sentinel-2 imagery was a good enough compromise that allows us to cover the entire season, whereas adapting input data to include CHIRPS monthly rainfall might result in either (a) lose of temporal resolution (S2 data at a monthly time-step) or (b) incorporated highly codependent rainfall data (CHIRPS rainfall at 10-day time-step). Nonetheless, we will perform an analysis of impact of such variables for classification in a simplified framework as will be explained in **Chapter 5**.

4.1.2 Unsupervised annual fallow mapping approach

As a complementary analysis, we implemented the procedure described in Tong et al. (2020) and use JECAM data as reference for testing the accuracy of the results obtained. Although this methodology has the advantage of not being constrained by a limited ground truth reference data set, which makes it easily upscalable to large extents, we keep in mind that its potential for producing accurate results in our case will be deemed compared to any supervised methodology which can learn from reference data, especially at such a small scale like the one referred here. However, we consider its assessment on our study site a fundamental step to further motivate our interest towards supervised classification approaches. For this reasons, we estimate that implementing Tong et al. (2020) unsupervised *fallow* mapping methodology to a single year is a sufficiently extended trial as for evaluating this approach. Year 2017 was chosen as the period to test this approach as this is the period where there is the largest amount of *fallow* fields registered in JECAM data set, allowing for a more suitable evaluation of generated maps against ground truth data.

This approach consist of two main steps, namely (a) the generation of a reference data set (b) an image classification step based on such reference (steps *a* and *b* from source flowchart in Tong et al. (2020), see **Fig. 4.2**). In both cases a Random Forest classifier is employed mainly differing in what data is used as reference and the input data used as predictor variables. The last step of the original workflow ("assessment") is not conducted in our implementation.

Reference data generation

This part of the approach is sub-divided into three steps which consist in (a) determination of optimal image selection period based on the detection of the mid- and end-of-season, performed using MODIS imagery, (b) generation of "first reference data" through unsupervised classification (clustering) using a Sentinel-2 composite over the previously identified period and (c) generation of a "second reference data" in which we train a Random Forest classifier using results from previous step as reference labels and produce a binary classification (cropped / fallow) of the entire cropland. Final reference data is

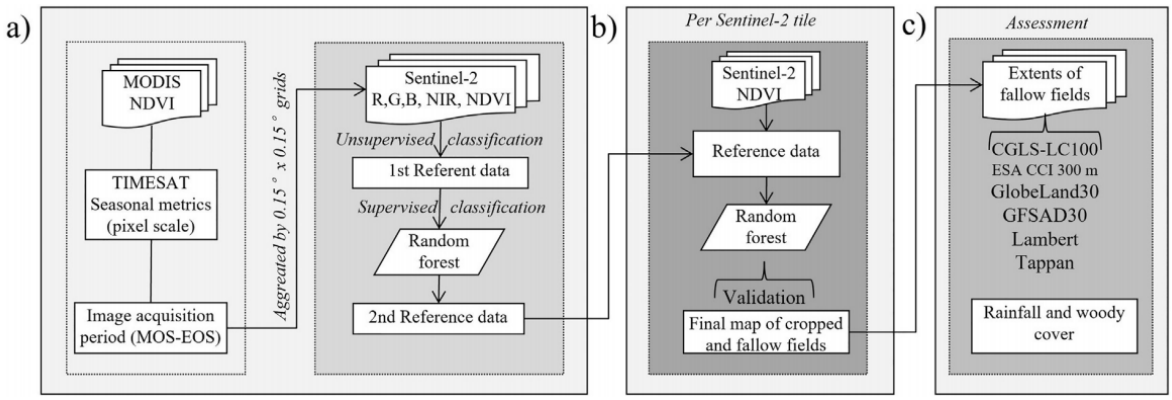


Figure 4.2: Original workflow schema of unsupervised *fallow* mapping. a) Generation of training reference data set, b) model training and test and c) assessment of the proportion of *fallow* land in different cropland products (Source: Tong et al. (2020)).

obtained by selecting aggregates of pixels most likely to approximate the shape of fields. This last selection step was customized for our case and further detailed are given in following subsections.

Optimal image selection period. This step is conducted by performing the fitting of MODIS 16-day composites NDVI band from year 2017 (MOD13Q1, v6) (Didan et al., 2015) to a double sigmoid, which allowing to represent a growing, plateau and senescent period. With this we can extract phenological information: mid of season and end of season specifically. This was conducted by means of OrfeoToolbox's (OTB) *Sigmofitting* application on version 7.2.0, taken from an official remote module of the toolbox (see Inglada (2017)). The optimal period was determined by the mid of season (MOS) and end of season (EOS) dates, which in turn are defined as the moment of 80% of amplitude before maximum NDVI and 50% of amplitude after maximum NDVI (Tong et al., 2020).

Generation of first reference data set. Sentinel-2 imagery is retrieved for year 2017 between obtained MOS and EOS dates. Median values over the temporal dimension for VNIR (red, green, blue and near-infrared) and NDVI bands are then computed to build a unique multi-band composite. A total of ten thousand pixels were randomly sampled within a cropland mask built from previous experiences on the study site using the *Moringa* processing chain (Gaetano et al., 2019). This sampling was then clustered using K-means algorithm to split data into two classes. The cluster whose centroid has the lowest NDVI value is considered as *cropped*, while the other (higher NDVI value) as *fallow* class respectively. Note that this choice stems from a fundamental observation made in the paper about the typical intra-seasonal dynamics of NDVI for fallowed and cropped fields in the Sahelian band, which is likely to be inappropriate for the sudanian site under study as observable from **Fig. 3.6**. However, since statistics of **Fig. 3.6** may be severely affected by the presence of clouds, we decided to go on with our test in order to definitely infirm the underlying hypothesis on NDVI for our study caseA. Finally, in

order to obtained highly contrasting classes, bottom 25% (below 25th percentile) and top 25% (above 75th percentile) pixels are selected for *cropped* and *fallow* class respectively based on the median NDVI values.

Generation of second reference data set. A second reference data set is then constructed by training a Random Forest classifier using as reference data the first reference data set and median VNIR and NDVI values as input data. Trained model is then employed to generate a classified map for the entire study site. Resulting map was employed to label the object of a segmentation issued from the median Sentinel-2 VNIR/NDVI image, which was generated using Baatz and Schäpe algorithm (Baatz et al., 2000) through the application available in the Orfeo Toolbox named *Generic Region Merging* (OTB v7.2.0, Inglada and Christophe (2009)). Segmented image parameters are set by trial and error to generate compact object of approximate field-like sizes and shapes. Object selection is done by filtering object with at least 75% class coverage, object area between 2-6 ha and a compactness index (perimeter/area) below 0.04. The selection procedure differs from original methodology, but in turn, yields more spatially accurate reference data as this process is handled following a more restrictive object-based approach.

Image classification and testing

Using filtered second reference data set as training set and the full Sentinel-2 NDVI time series as input predictor variables, we trained a Random Forest model and classified the entire cropland into cropped or fallow. The resulting map is then tested using the original JECAM's data set.

4.2 Results

4.2.1 Supervised annual fallow mapping approach

For this approach multiple we trained multiple Random Forest (RF) and TempCNN (TCNN) image classifiers using the standard 5-fold cross-validation procedure as described in **Sec. 2.3.3** on an annual basis. As said, the classification task is divided into six target classes, including a *fallow* class. Trained models resulted into an average overall score (OA) of respectively 89% and 82% for Random Forest classifier (RF) and TempCNN for the period 2016-2021 (year 2019 not included). Average user accuracy (UA) was 81% and 70% for RF and TempCNN respectively for the same period (2016-2021), whereas producer accuracy (PA) were 70% and 79%. An extract of accuracy metrics is presented in **Tab. 4.1**. For more detailed results see the full table **Tab. A.1**. The F1-score is here referred as a more adequate metric to evaluate the specific detection of the *fallow* class, given the highly unbalanced nature of the classification task. Nonetheless, when analyzing class-wise F1-scores (see **Fig. 4.3**) *fallow* class is noticeably low compared to the rest of

other classes considered, with best results obtained with TempCNN trials yet below the 0.40 mark.

Looking at the maps generated with the corresponding models, the RF trial (see **Fig. 4.4**) did not manage to detect *fallow* land outside the few instances that correspond to locations where reference data (used for training) was available, hence showing an insufficient generalization capability for this class. In contrast, trials where TempCNN is employed (see **Fig 4.5**) resulted into slightly better results than RF, which translated, in turn, into a far more significant presence of *fallow* class (in red on the maps). For the examples shown, the proportion of *fallow* class over total cropland surface (i.e. sum of *cropped* and *fallow* classes) reached 0.03% and 38.60 % for RF and TCNN trial respectively, with values in opposite extremes and both significantly disagreeing with *fallow* estimates in JECAM data set (4% of cropland sampled).

Table 4.1: Supervised fallow annual mapping accuracy metrics. UA, PA and OA stand for user's accuracy, producer's accuracy and overall accuracy respectively. Trial classes RF, TCNN stand for Random Forest and TempCNN image classification methods respectively.

Trial	Year	OA	Class	UA	PA
RF	2016	0.87	cropped	0.85	0.98
			fallow	0.29	0.01
RF	2017	0.89	cropped	0.91	0.98
			fallow	0.03	0.00
RF	2018	0.90	cropped	0.90	0.97
			fallow	0.20	0.00
RF	2020	0.88	cropped	0.92	0.98
			fallow	0.11	0.01
RF	2021	0.91	cropped	0.92	0.98
			fallow	0.92	0.10
TCNN	2016	0.75	cropped	0.94	0.66
			fallow	0.07	0.40
TCNN	2017	0.78	cropped	0.97	0.75
			fallow	0.11	0.71
TCNN	2018	0.83	cropped	0.94	0.83
			fallow	0.06	0.36
TCNN	2020	0.85	cropped	0.96	0.92
			fallow	0.19	0.70
TCNN	2021	0.89	cropped	0.96	0.91
			fallow	0.18	0.63

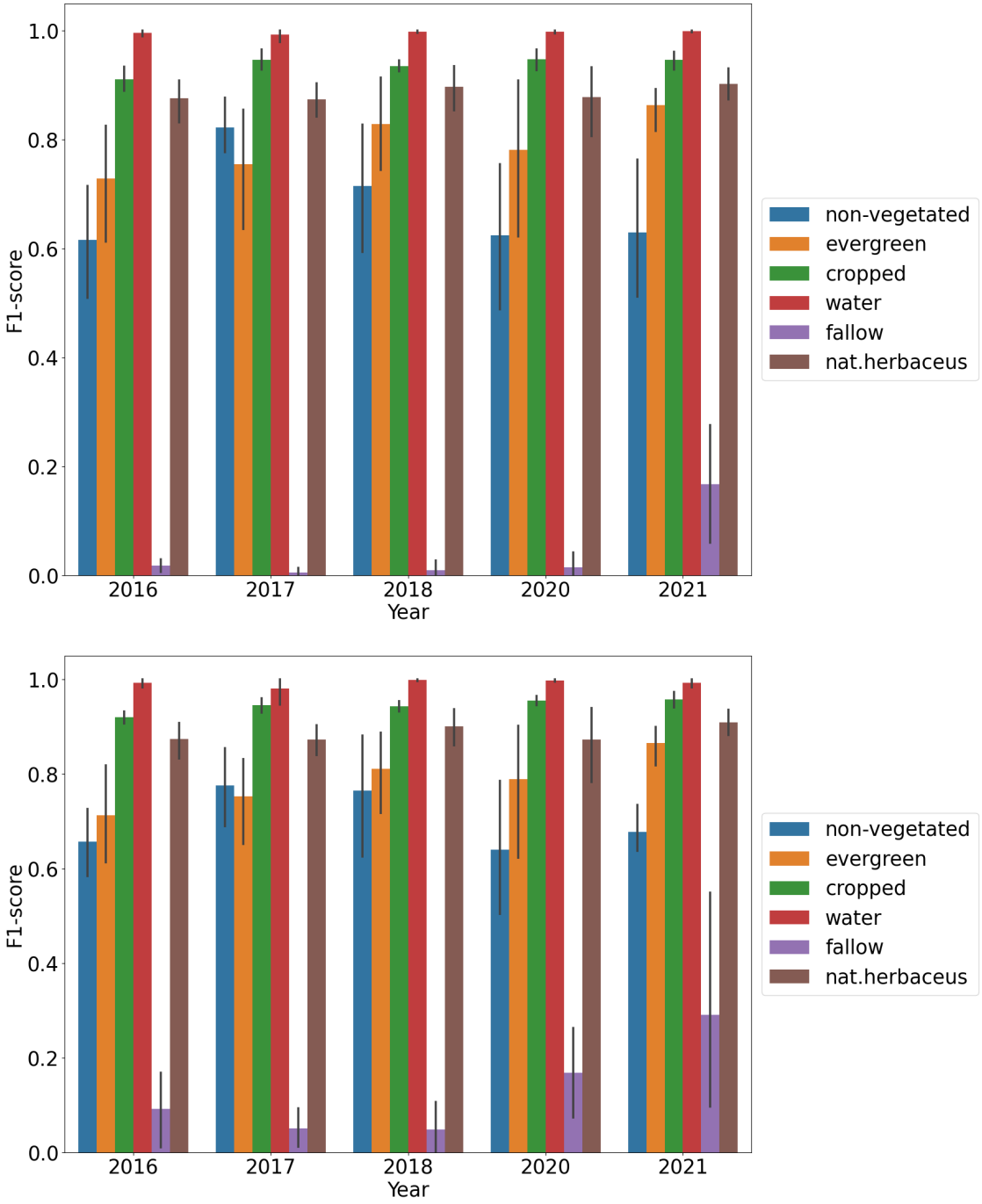


Figure 4.3: F1-scores for supervised annual fallow mapping strategies using Random Forest (top) and TempCNN (bottom) classifiers. All trials were tested using JECAM's data as reference data set.

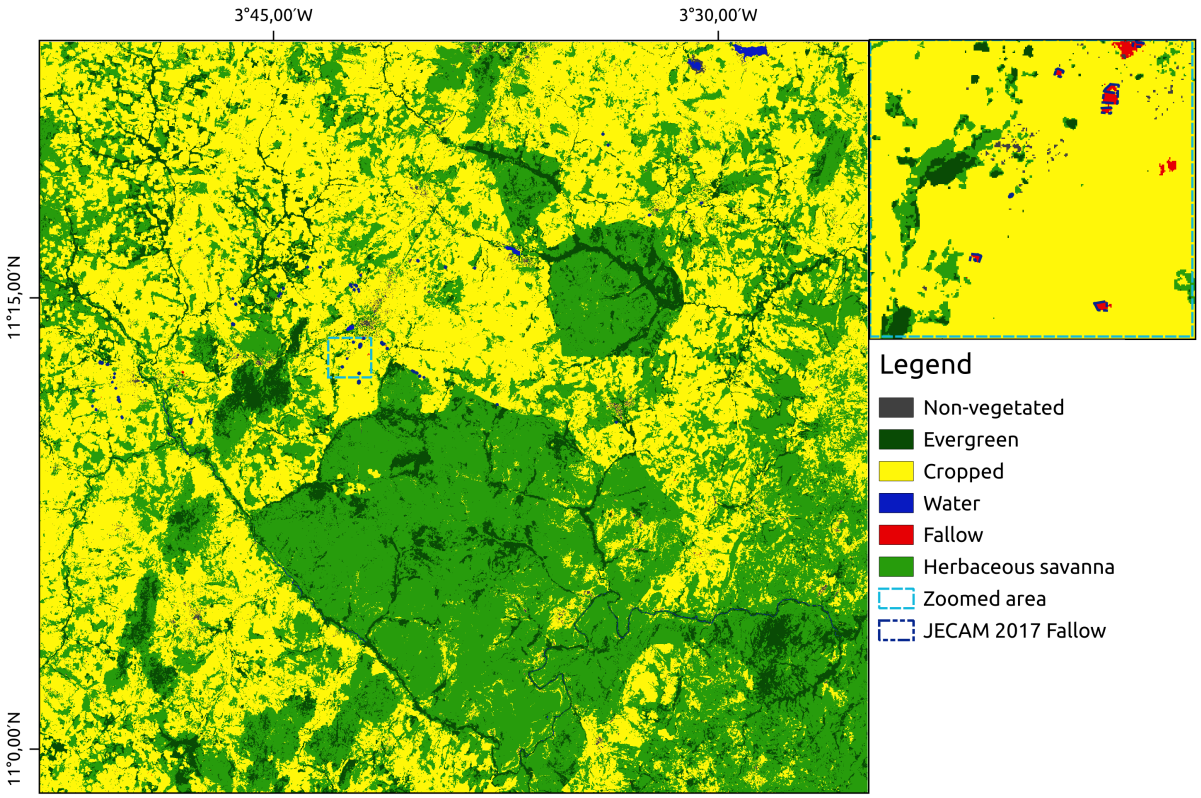


Figure 4.4: Example of classification results for year 2017 for RF trial. Top right zoom-in shows area with known fallow fields as registered in ground truth data.

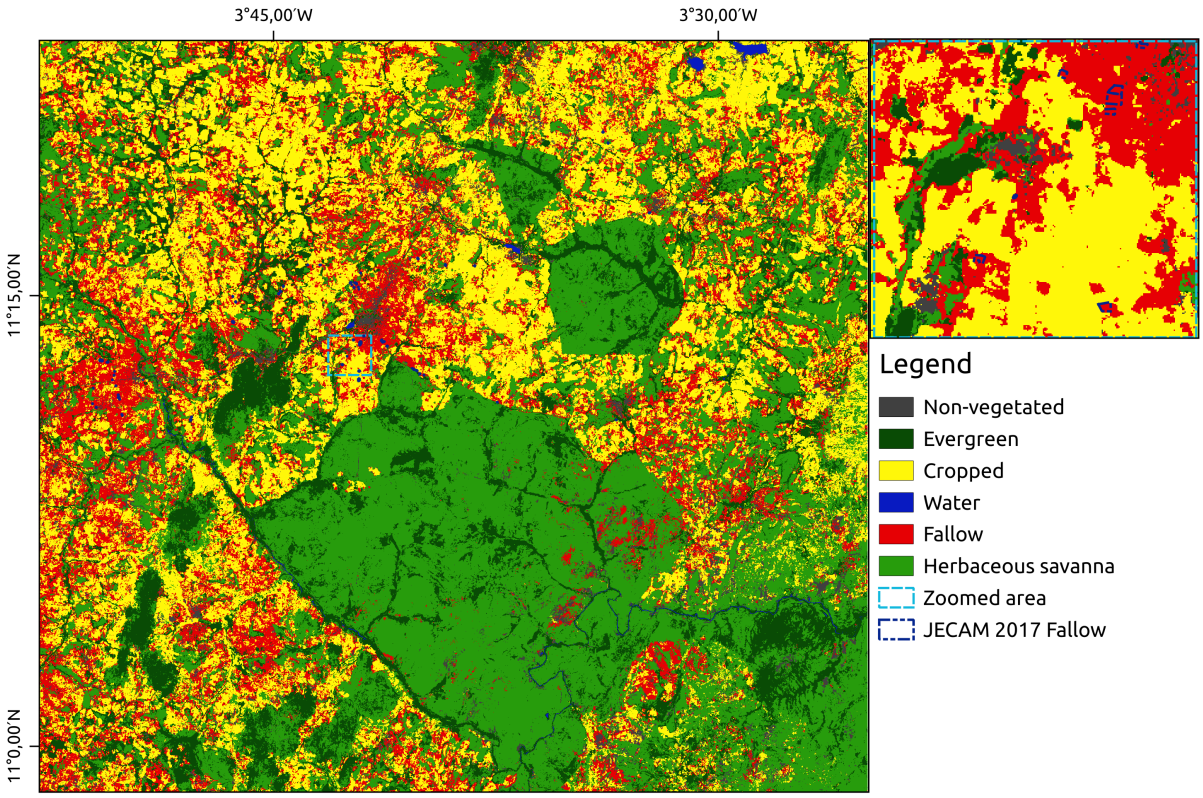


Figure 4.5: Example of classification results for year 2017 for TCNN trial. Top right zoom-in shows area with known fallow fields as registered in ground truth data.

4.2.2 Unsupervised annual fallow mapping approach

Reference data generation

First of all, we show the final second reference data set object from the unsupervised methodology are shown in **Fig. 4.6**. Resulting object are scattered across the cropland area of our study site and yielded a total amount of 546 *cropped* objects and 256 *fallow* objects. These objects are independent from those contained in JECAM data set, hence the number of available data points (pixels) as well as the number of classes considered differs significantly when compared with JECAM's 2017 data.

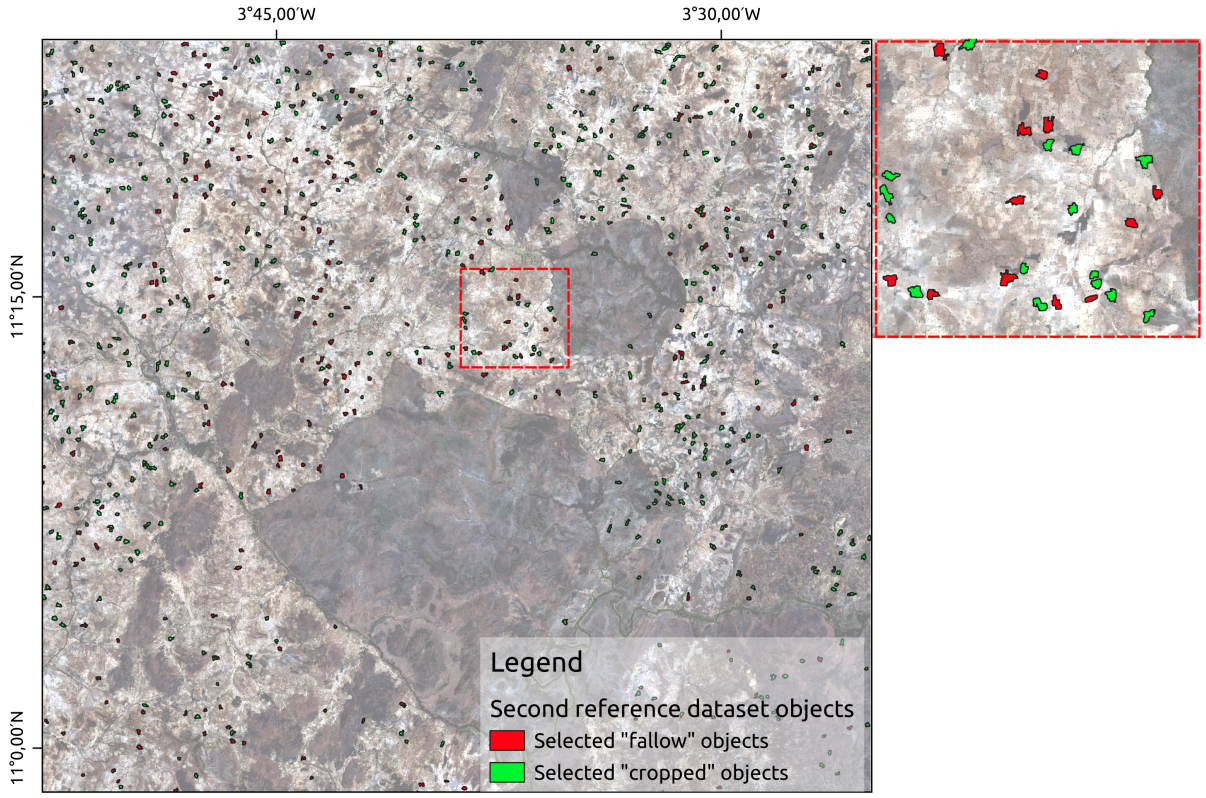


Figure 4.6: Filtered second reference data set obtained for Semi-supervised approach for year 2017. Selection of objects followed size and compactness criteria.

The spatial distribution of this data set covers a larger portion of the study site. This is an expectable result due to its unsupervised-based sampling, where the only limiting spatial factor is the cropland mask employed. By contrast, although highly more accurate, the JECAM data set presents a more clustered data, that locates in the vicinity of main roads. This is due to the sampling protocol it self used in field missions that compose JECAM data, causing more isolated and less accessible fields not to be included.

Classification results

Once the reference data set is built accordingly to procedure described in **Sec. 4.1.2**, we could train a Random Forest model using the selected compact objects from the second reference data set (see **Sec. 4.2.2**) as model's input labelled reference data and

the Sentinel-2 extracted NDVI time-series for the year 2017 as input features. To evaluate and calculate model accuracy metrics we used the now trained model to predict the classes *cropped* and *fallow* for the entire study site and evaluated predicted results using 2017 JECAM's Koumbia data set. Final unsupervised classification approach evaluated against JECAM ground truth data are present in **Tab. 4.2**.

Table 4.2: Unsupervised annual fallow mapping confusion matrix (in number of pixels) and accuracy metrics. UA, PA and OA stand for user's accuracy, producer's accuracy and overall accuracy respectively. Trial class UN stands for *unsupervised* classification approach. C and F stand for cropped and fallow, whereas subscripts r and p stand for reference and predicted in confusion matrix.

Trial	Year	OA	Class	C_r	F_r	UA	PA	F1-score
UN	2017	0.44	C_p	9662	517	0.95	0.44	0.61
			F_p	12052	287	0.02	0.36	0.44

Results in **Tab. 4.2** show a high level of both commission and omission error towards *fallow* class, with more than 50% of reference *cropped* pixels miss-classified as *fallow* and more than 60% of reference *fallow* pixels miss-classified as *cropped* class, represented by a low user accuracy and producer accuracy respectively. The *cropped* class is not as severely affected in its user accuracy as a consequence of the reduced amount of reference *fallow* pixels, hence with the effect to further increase commission error (i.e. reference *fallow* miss-labelled as *cropped* class). Note that in this approach, as a consequence of the use of a cropland mask, the amount of reference data points (i.e. Rasterized JECAM 10 meter pixels) is significantly reduced, with a total of 21714 for *cropped* and 804 *fallow* class pixels available out of over 50000 and 2000 pixels in source JECAM data set for year 2017 (see **Fig.4.1**). Needless to say, overall model ability to discriminate the targeted classes is subjected to the quality of the cropland mask employed during the training process. If we attend to F1-score values as a single measure that can represent the per-class overall accuracy, the proposed method exhibit a clearly poor performance for *fallow* class in our case.

In **Fig. 4.7** we present the results of the predicted map over the entire study site. It can be noticed a high level of overestimation of the *fallow* class across the study site, accompanied with general salt-and-pepper artifacts. Assuming that the proportion of *fallow* land in JECAM during 2017 is sufficiently representative of the entirety of the study site, we can estimate that c.a. 4% of cropland in JECAM data set is considered as *fallow* compared to around 42% for this trial. Moreover, it can be noticed that most of the *fallow*-annotated JECAM fields are miss-labelled as *cropped* class (see zoomed image in **Fig. 4.7**). As a further evidence of the inappropriateness of the provided map, in no case we can observe isolated pixel aggregates exhibiting the typical field size or shape and labelled as *fallow*.

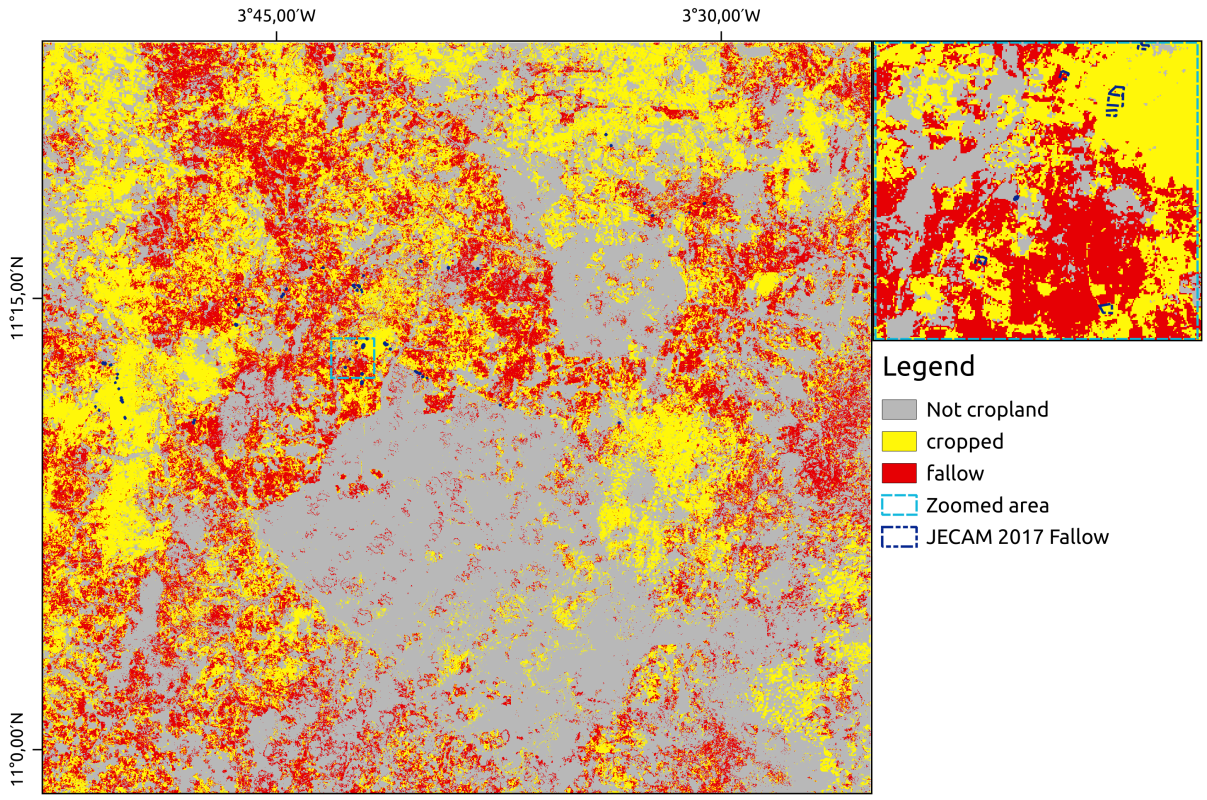


Figure 4.7: Unsupervised annual fallow mapping image classification result. Notice that this trial is a binary classification, grey color was added to enhance contrast. Top right zoom-in shows area with known fallow fields as registered in ground truth data (dark blue dash line fields).

4.3 Conclusions

In this chapter we tested an unsupervised annual fallow mapping approach based on Tong et al. (2020) procedure and a supervised annual fallow mapping approach. In both strategies the final discrimination is performed by an automatic image classifier trained with Sentinel-2 time-series and where *fallow* land is considered as another cropland subclass. Results were evaluated against JECAM ground truth data, showing that these methodologies are not sufficient to provide any viable estimate of *fallow* land. Resulting model showed a lack of discriminative power indicating that methodologies employed cannot capture the potential explanatory information contained in radiometric features.

Chapter 5

Trajectory analysis

On this chapter we present a different methodological approach for mapping the proportion of cropland that is not being actively cropped. This methodology addresses some of the limitations related to the direct mapping of fallow land as a specific cropping practice, by relying on an original land trajectory modelling which exploits the multi-year nature of the reference data. This chapter made the basis of a scientific publication (see Castro Alvarado et al. (2023)).

5.1 Non active agricultural land mapping approach

The analysis and experimental results presented so far in this thesis have given quite an evidence of the challenge that represents the mapping of *fallow* practice using satellite image time series. On one hand, they seem to prove that the seasonal dynamic of vegetation, as measured by Sentinel-2 imagery, does not contain enough information to distinguish fallow land from any other vegetated class with an annual vegetation cycle (either agricultural land uses or natural deciduous vegetation). On the other hand, no external sources of information among those tested have proven effective as a spatial proxy of the presence of fallow land, which could be effectively used as prior knowledge in an automatic remote sensing based mapping process.

For these reasons, the idea here is to somehow reformulate the mapping problem, by tackling the identification of non-active agricultural land (NAAL) as a whole, rather than the fallow land intended as previously active cropland undergoing an explicitly sought regeneration process prior to a new cropping period. NAAL is here more loosely understood as known cropland that is no longer cropped and with *unconfirmed recultivation*. This means that it includes the *stricto sensu* fallow land as mentioned, but it also encompass all other types of "abandoned" cropland, and therefore ends up playing a similar role w.r.t. fallows in the context of the agro-system.

To be more specific on the proposed approach, rather than attempting to differentiate NAAL from other types of uses based on annual land cover data, we start from the observation that a way to trigger the identification of NAAL is to detect, over multiple cropping

seasons and over areas exhibiting seasonal vegetation life cycle (cropland-like), transitions between periods in which vegetation growth is *managed*, i.e. a cropping activity is actually performed, and periods in which it exhibits an *unmanaged* condition, which may happen at seasonal scale for both NAAL and areas with natural deciduous/herbaceous vegetation. Based on the hypothesis that an annual time series of optical multi-spectral images may enable the accurate identification of these two states (managed / unmanaged vegetation over a cropping season), our approach relies on such annual LULC mapping and the subsequent analysis of these maps over multiple years to provide a consistent, field-level accurate mapping of NAAL.

Obviously, the main constraint of this approach is the availability of a proper reference data set covering multiple consecutive years, which provides information, for each year and with no ambiguities, on whether a labeled surface belongs to the active cropland, and hence to a *managed* agricultural land class, or to an *unmanaged* class, which can in turn include both non active cropland and natural herbaceous vegetation. Information might also be provided for land cover classes with more stable dynamics across time, information which can be useful to detect and mask out such areas which are less related to the purpose of NAAL mapping. Needing no specific details on these classes, we could complete our annual land cover model by grouping them all under two classes, namely *evergreen* and *non-vegetated*, to address respectively the constant presence and absence of green vegetation.

As already mentioned in the data description section, the JECAM data set provides annual information on both land cover and crop group classes, and it can easily be used to comply to the 4-class model (*managed*, *unmanaged*, *evergreen*, *non-vegetated*) and allow the production of the annual base maps used for multi-year NAAL identification.

On these premises, the method we intend to propose here is quite straightforward, composed of three stages tackling respectively : (a) the production of annual base maps inferring the 4-class land cover model to the entire area, (b) the application of transition rules for the analysis of land cover trajectories allowing the identification of NAAL, and (c) a suitable processing of the original reference data set to provide a reliable and fair validation of NAAL detection. An overall scheme of the proposed three-stage method is depicted in **Fig. 5.1**, a detailed description of each processing block is provided in the following subsections.

5.1.1 Annual management mapping

Our first step consists in training a set of annual supervised classification models using the corresponding time series of Sentinel-2 images as a main input. A first straightforward possibility could have been to simply perform a supervised classification directly using the JECAM data set as-is to train the classifier. However, it is clear from **Tab. 2.1** and the results presented in **Chapter 4**, that a strong class imbalance is present in terms of annotated surface, which may lead to poor performances on less represented classes.

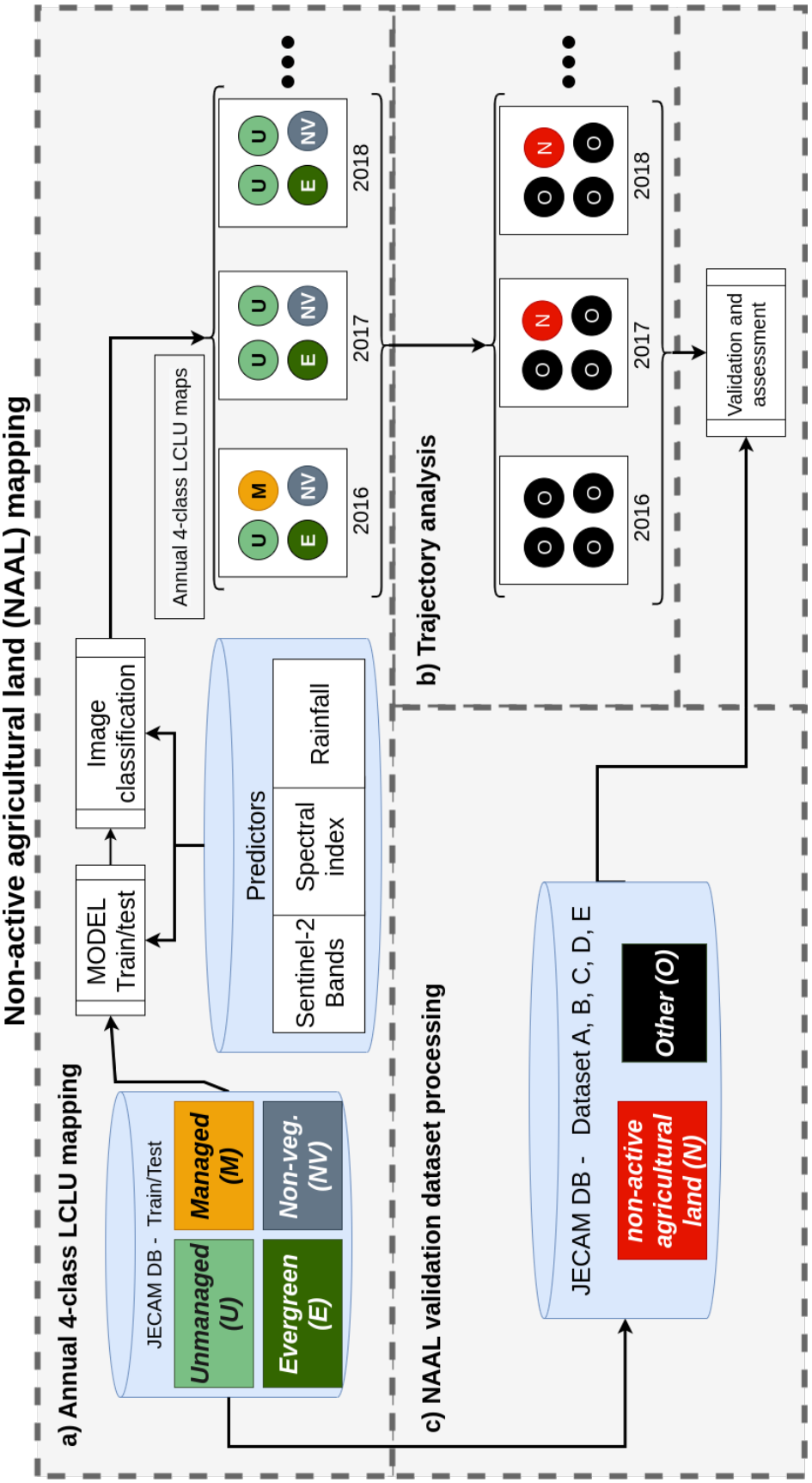


Figure 5.1: NAAL mapping general workflow: (a) annual land use mapping and generation of 4-class maps; (b) Trajectory analysis of land use maps and NAAL detection; (c) Reference data set processing for the validation of the NAAL maps.

This is particularly true for the *fallow* class which, as arguably deductible observing the entangled NDVI profiles shown in **Fig.3.6** (**Chapter 3**) (for year 2017, without loss of generality), is likely to be confused with some active crop class and hence erroneously detected as *managed*, in contradiction to the provided definition of NAAL.

Hence, we here decide to pre-process the JECAM data set prior to the supervised classification process, by relabeling samples of the active crop classes into a single *managed* class (namely the *cash crops, cereals, leguminous and oilseed* classes), and grouping *fallows* with other deciduous vegetation classes (namely *herbaceous, shrubby savanna and pastures*) into the *unmanaged* class. *Woody savanna* and *forest* fall into the *evergreen* class, while the *non-vegetated* class includes all constantly non-vegetated areas (*bare soil, built-up, water bodies*). We report in **Fig. 5.2** the final number of pixels per class after the proposed class grouping and the rasterization of the source vector layer using the Sentinel-2 10 m grid. As expected, the final result shows a much more balanced configuration of class samples, especially between the *managed* and *unmanaged* classes whose transitions through subsequent years are susceptible to trigger the detection of NAAL areas. The overall classification problem is simplified, with the least represented classes being the *non-vegetated* and *evergreen* areas which are in turn the easiest to discriminate using radiometric information.

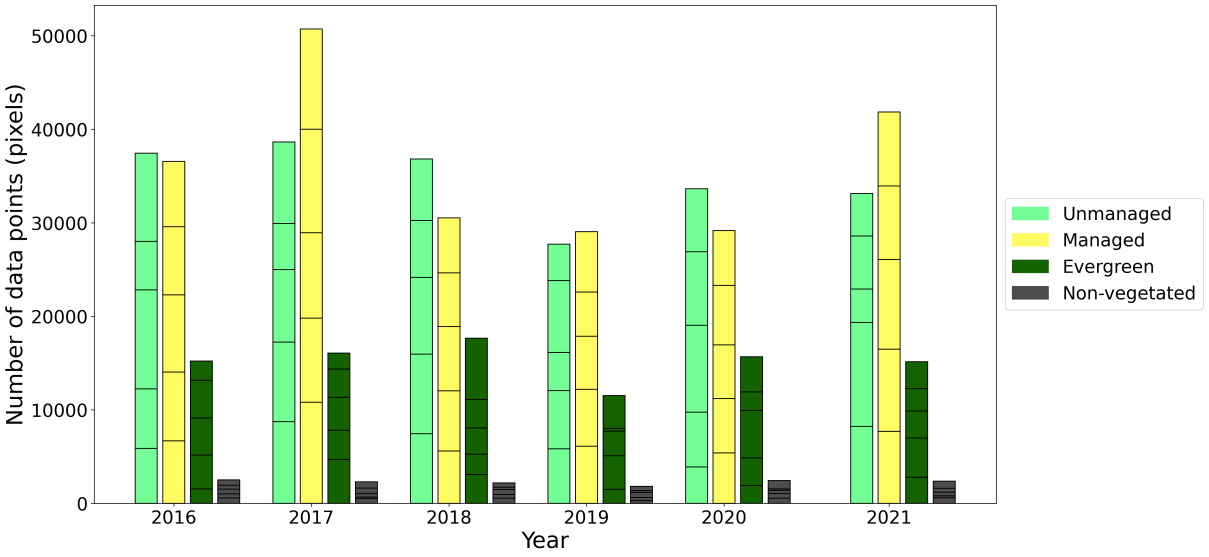


Figure 5.2: Number of data points in pixels (after rasterization using the Sentinel-2 10-m grid as spatial reference) for all years and separated by management land use classes. Bar splits represent the proportion of pixels employed in each cross validation fold, see **Sec. 5.2.1**.

Following an ensemble classification strategy based on the Random Forest algorithm, we decide to use the whole gap-filled time series, along with the derived indices (vegetation-related and soil) as well as rainfall data as predictor variables. A detailed description of the experimental setting of this phase along with an insight on the importance of such variables is provided in the experimental section of this chapter. In parallel we conducted a similar classification strategy implementing TempCNN algorithm for exploiting possible temporal

correlation of input variables during the season, but using exclusively the Sentinel-2 image time series (and the relative indices), for the reasons already mentioned in **Sec. 4.1.1**.

5.1.2 NAAL detection

Annual 4-class maps provided in previous step over the available time span, say X_t with $t \in \{1, \ldots, n\}$, are then employed for the analysis of trajectories implying classes *managed* (M) and *unmanaged* (U) at pixel level to derive the set of $n-1$ NAAL (N) maps Y_t over $t \in \{2, \ldots, n\}$ (as a consequence of our causal inter-annual analysis, NAAL identification for the first year of available reference data can not be computed). Our trajectory hypothesis is quite simple, and based on the following two main rules (see the right-upper part of **Fig. 5.1**):

1. any transition from *managed* to *unmanaged* over two subsequent years on a pixel in a location (u, v), say $X_{i-1}(u, v) = M$ and $X_i(u, v) = U$, implies "flagging" a pixel as NAAL on year i, say $Y_i(u, v) = N$;

2. if a pixel is flagged as NAAL for a given year, any subsequent *unmanaged* label on the same pixel is automatically classified as NAAL on the output to maintain temporal and spatial consistency, say:

$$Y_i(u, v) = N, X_j(u, v)|_{j \in i+1,\ldots,n} = U \Rightarrow Y_j(u, v) = N$$

In order to reduce possible ambiguities, and on the hypothesis that classes which are less reliant on vegetation dynamics are usually more accurately discriminated, occurrences of either the *evergreen* and *non-vegetated* class on a pixel over time lead to the automatic classification of this pixel as non-NAAL (O) for the whole period. This is done under the implicit hypothesis that no significant changes (such as artificialization or tree plantations) have taken place over the period.

5.1.3 NAAL validation and assessment

In the proposed NAAL classification workflow, a major issue is represented by the necessity to work out a test reference data set to validate the methodology and assess the accuracy of the NAAL maps output by the multi-year trajectory analysis. Matter of facts, we do not dispose of a multi-year data base of properly annotated NAAL fields. In principle, we could limit our validation to the original JECAM data set and consider the *fallow* class for NAAL assessment. Yet, only relying on this would be unfair, since much more of the arable land could be *de facto* included into the wider NAAL class, including a part of the fallow land which has been annotated as belonging to another deciduous vegetation class because of missing explicit information on cropping practices or abandoned agricultural fields.

However, it is possible to derive some NAAL-specific validation set from the original JECAM data set by applying the same rules that we use for trajectory analysis on land cover maps, using objects from the data base instead of map pixels, and the intersection of objects over multiple years to track transitions between the *managed* and *unmanaged* classes. This is made possible by the fact that, as already mentioned in **Chapter 2**, the field survey protocol for the Koumbia site was enforced by taking annotations over the same fields year-by-year to the extent possible. Still, this process is more subtle than it may appear, since a sufficiently large quantity of surface which is overlapping over time and may be labeled as NAAL (which appears to be a small percentage of the area according to the JECAM data set) is in fact needed. Moreover, field boundaries are not static across years, which may lead to a dramatic reduction of usable surface if a simple intersection rule is applied, and in particular a single missing annotation for a given field will make all the data from the relative field unusable.

Thus, we decided to provide several versions of the validation data set, each obtained with a different set of rules for the analysis of the overlaps over time. This goes from a data set obtained by a rigorous intersection of polygons, which is very precise in term of NAAL identification but with a possibly insufficient surface for a reliable assessment, to progressively admitting extrapolation of missing data in space and time, in order to extend the total validation surface, and hence the statistical sufficiency, at the price of some "noise" in the resulting data set.

The different processing approaches over the reference data set, rendered so as to comply to our 4-class model for trajectory analysis, are described below and summarized on **Fig. 5.3**. All these processes start with a rasterization of the reference polygons, available in vector format, using the Sentinel-2 grid as geometric reference. As already mentioned, all proposed strategies aim at the relabeling of the whole set of reference pixels from the 4-class nomenclature used for annual mapping to a binary (NAAL vs. non-NAAL) nomenclature based on the same transition rules applied for trajectory analysis. This means that all reference pixels that, for a given year, are not concerned by a relabeling to NAAL are grouped in the unique antagonist *non-NAAL* class. Obviously, the resulting data sets are expected to be strongly unbalanced (NAAL accounts only for a small percentage of the whole area), but in this case the unbiased assessment of the detection of NAAL areas will eventually be possible using per-class figures (*precision*, *recall* and *F1-score*).

Data set A : Original JECAM data set This validation set is simply made up of the original JECAM Koumbia data set in which the objects are relabeled to fit the 4-class trajectory model. No spatial or temporal modification is applied on the source data. This represents the original data set where only "fallow" fields are considered as NAAL data. There is no data alteration on this case.

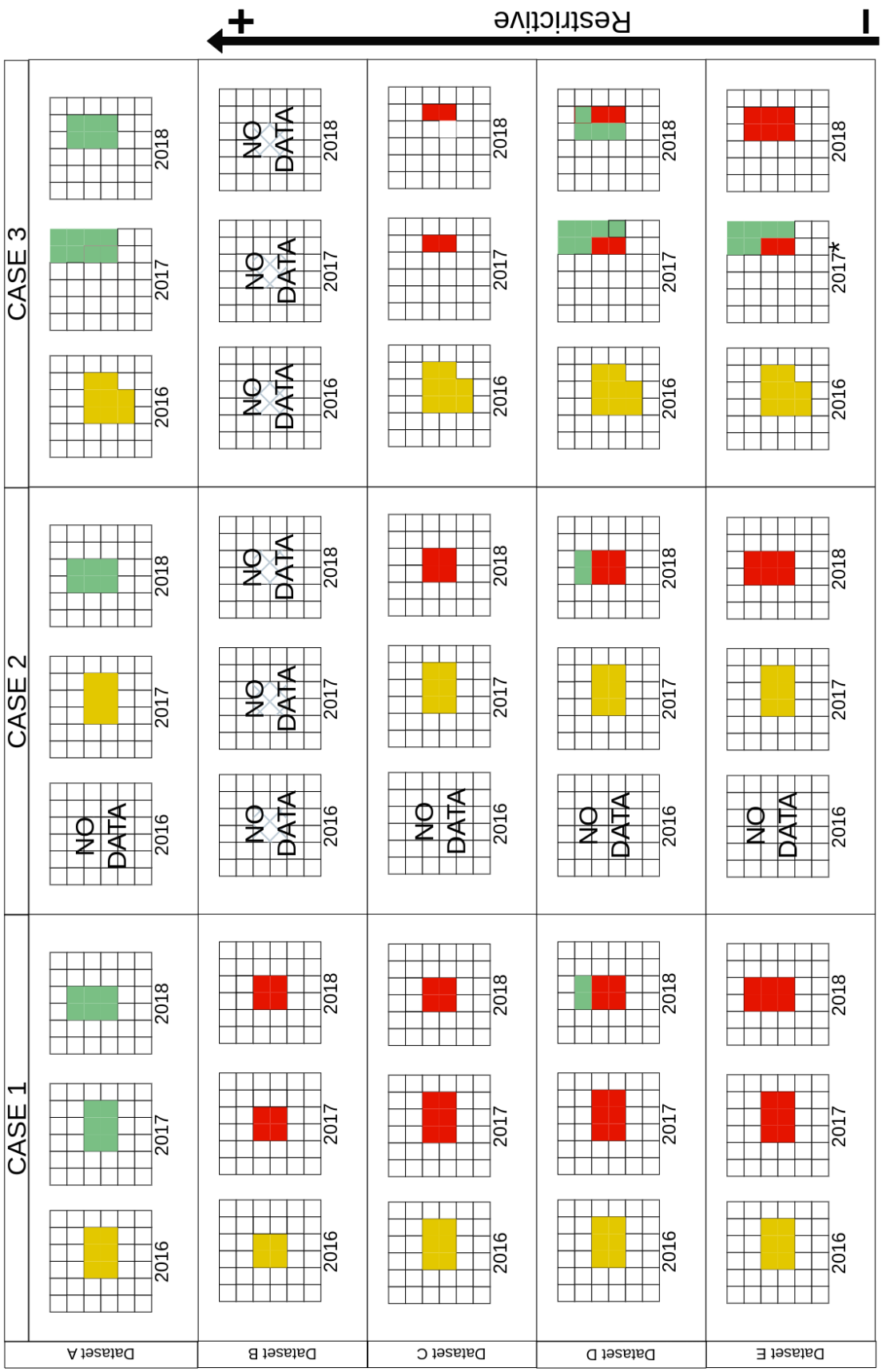


Figure 5.3: Illustrative scheme of validation data set processing through 3 examples over 3 consecutive years: Case 1 with no missing years, Case 2 with first year missing, and Case 3 with no stable field boundaries across years considered. Data set A corresponds to the rasterized version of the original JECAM data set, relabeled to fit to the 4-class model (*managed, unmanaged, evergreen, non-vegetated*).

Data set B : Strictly reliable surfaces This is the case when the rigorous multi-year intersection of reference objects is performed, basically retaining only the most reliable surfaces. In other words, only those pixels where data is available in all consecutive years are considered for trajectory analysis and NAAL relabeling. Any point with missing data for a given year is dropped out of the data set.

Data set C : Locally reliable surfaces In this scenario, we make the implicit assumption that the validity of a surface for a given year only depends on its past validity. This means that, for each year, the intersection is only performed with the surface of the overlapping object in the previous year. In this way, all pairs of pixels from two consecutive years that are eligible to be relabeled to NAAL are taken into consideration, regardless of what happened before or after, while the portion of object which is non comparable is discarded. Compared to *data set B*, this data set do not guarantee that a comparable set of surfaces are used for validation year-by-year, leading to possible discrepancies in the multi-year assessment strategy.

Data set D : Sub-field NAAL relabeling Here, we basically propose the same logic than for *data set C* at pixel level, with the exception that the portions of objects which are left out from an intersection between two consecutive years are not discarded, and can still be used for NAAL relabeling in subsequent years. Alternatively said, relabeling can occur at sub-field level.

Data set E : field-level NAAL extrapolation Based on the observation that field boundaries can move through time, and considering how densely cultivated the area is, for this data set we make the underlying hypothesis that, if a certain field is known to have been actively cropped on year Y_i, a newly annotated *unmanaged* field on $Y_i + 1$ whose surface is only partly overlapping with the past year's field is probably a whole NAAL field. Hence, similarly to *data set D*, in this case we still consider pairwise overlaps between consecutive years and no data is lost from the source objects after processing. However, relabeling to NAAL can be propagated to an entire object if the intersection of the given object with the one from the previous year is larger than a given threshold in terms of area percentage, here heuristically fixed to 30%.

A resume of the total number of NAAL pixels for each validation data set and per-year is reported in **Fig. 5.4**, along with the percentage it represents with respect to the full data set (including non-NAAL pixels). Note that, by only retaining the strictly reliable surfaces, data set B systematically provide a severe selection of NAAL pixels w.r.t. the total samples, but the other strategies which try to extrapolate NAAL areas at the plot scale progressively tend to first restore the original amount and ratio observable in the data set A (actual *Fallow* fields w.r.t. the total annotated surface), then add some potential NAAL surface in reasonable measure (within a factor of 2). The slight drop in the amount of NAAL pixels in 2020-21 is due to the absence of *Fallow* annotated polygons

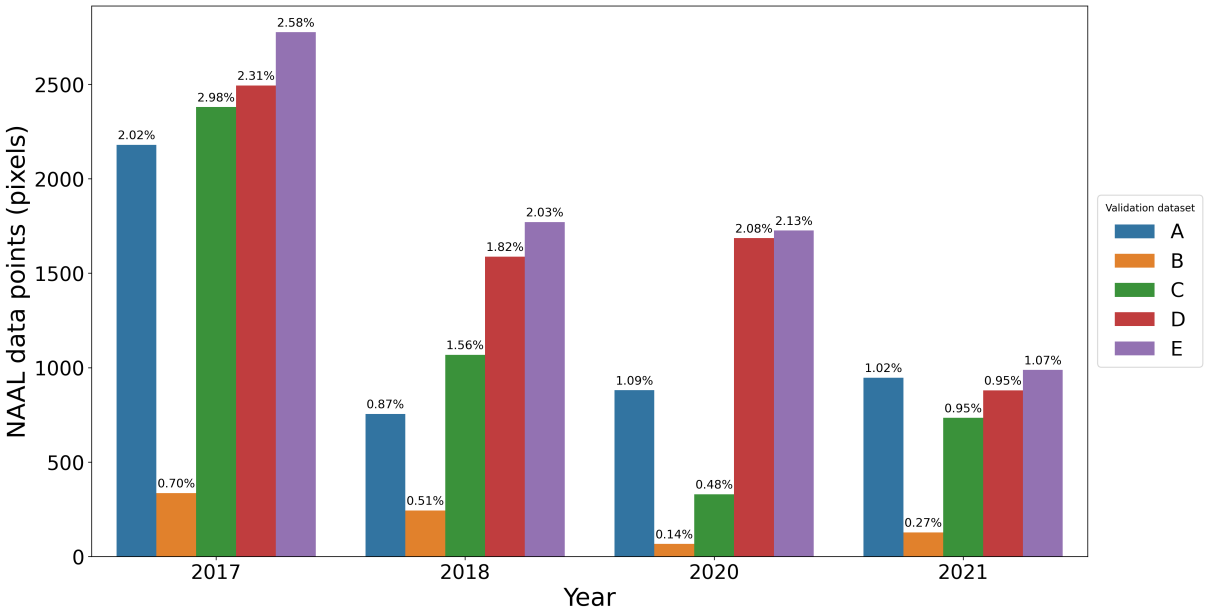


Figure 5.4: Total number of NAAL Sentinel-2 pixels in the validation data sets before (A) and after (B to E) processing. The validation data set A corresponds to the JECAM original data set in which the NAAL data pixels correspond to the "Fallow" polygons. For the record, only the cropped fields were registered for year 2019. Values on top of bars show the proportion of NAAL pixels over the total amount of available pixels for that given validation data set and year.

in the original JECAM data base in 2019.

5.2 Implementation of the approach

5.2.1 Annual LULC maps

For the production of annual 4-class land use maps reliable enough to enable trajectory analysis for NAAL detection, we carried out a strategy based on multiple per-year Random Forest (RF) and TempCNN (TCNN) classifications. Such solution deemed necessary to cope with the relatively low quantity of available annotated surface, especially if compared to the complexity of the landscape and the dimensionality of the input variables (593 and 544 variables per pixel for in total for RF and TCNN trials respectively), as well as to ensure comparability among the different maps.

We implemented the standard 5-fold model training procedure described in **Chapter 2** and make use of a 4-class based management nomenclature as described in **Sec. 5.1.1**. Resulting data splits using this nomenclature are shown in **Fig. 5.2**, where due to the splitting occurring at the polygon level sample selection translates into numbers of reference pixels in per fold/class/year. We used the standard parameterization described in **Chapter 2** to train our models and then generated an annual 4-class map by averaging the probability maps of the 5 models and labeling pixels according to the class with the highest probability. Variable importance (averaged over folds as well and only available

for RF models) was also computed for each year in order to identify the main predictors and check for consistency across years. Additionally, maximum per-pixel class probability was used as a proxy for identifying zones of lower model certainty.

5.2.2 NAAL mapping

Once 4-class maps were generated we perform the trajectory analysis as described in **Sec. 5.1.2** via custom scripts written on Python 3.8.

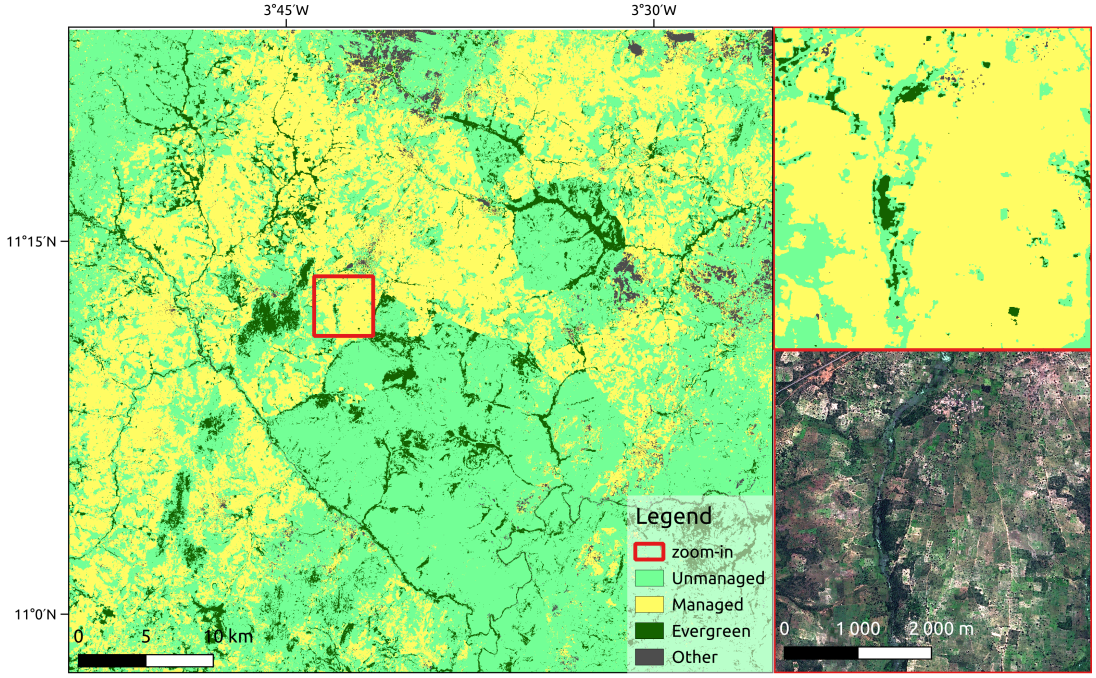
5.3 Results

5.3.1 Annual management mapping

The Random Forest and TempCNN trained models were employed for generating 4-class land use maps for the entire study site (see example in **Fig. 5.5** for 2017). The average overall accuracy (OA) across all years considered (2016-2021) was of 91.4% (RF) and 91.2% (TCNN), whereas the average user accuracy (UA) was 91.7% (RF), 89.8% (TCNN) and producer accuracy (PA) was 87.3% (RF), 86.9% (TCNN) all classes considered during the same period. Full report of map accuracies is provided in Annexes (see **Tab A.2**). Classwise minimum F1-scores values for RF trials were 0.91, 0.86, 0.80 and 0.75 for *managed*, *unmanaged*, *evergreen* and *non-vegetated* classes, whereas for TCNN trial these values were 0.93, 0.86, 0.76 and 0.75 respectively. For both algorithms *managed* and *unmanaged* classes were consistently the classes with the best average scores across all the considered period (**Fig. 5.6**), except for year 2019 with TCNN.

Given the globally accurate results provided by these classification tasks, it is now possible to perform an in-depth analysis of the importance of variables using the RF models. Per-year RF's variable importance averaged across models is shown in **Fig. 5.7**. In order to simplify the reading of such list, the importance for all multi-temporal variables has been averaged over time (e.g. B4 is the average importance of every 10-day Sentinel-2 red bands), but the maximum value of these variables over time is also depicted in red. We here observe that the "hierarchy" of such importance is globally stable over the years, with bands from the red and SWIR spectrum, along with some vegetation indices (NDVI, NDRE) steadily occupying the highest places in the list. Noticeably, both soil indices (RI and CI in particular) and rainfall data have a relatively high importance compared to the large number of spectral variables. Of course, maximum values show that if we consider individual time-steps for each variable, the absolute importance is higher for several spectral bands and indices, but still both external sources of information have a significant and systematic impact in all years' model decisions. For rainfall data in particular, which varied significantly across years considered (see **Fig. 2.2**), importance results support the hypothesis that even at low resolution such information may impact the quality of annual mapping of *managed/unmanaged* surfaces, independently of inter-seasonal variations. Yet, this did not translate into noticeable differences between RF

(a)



(b)

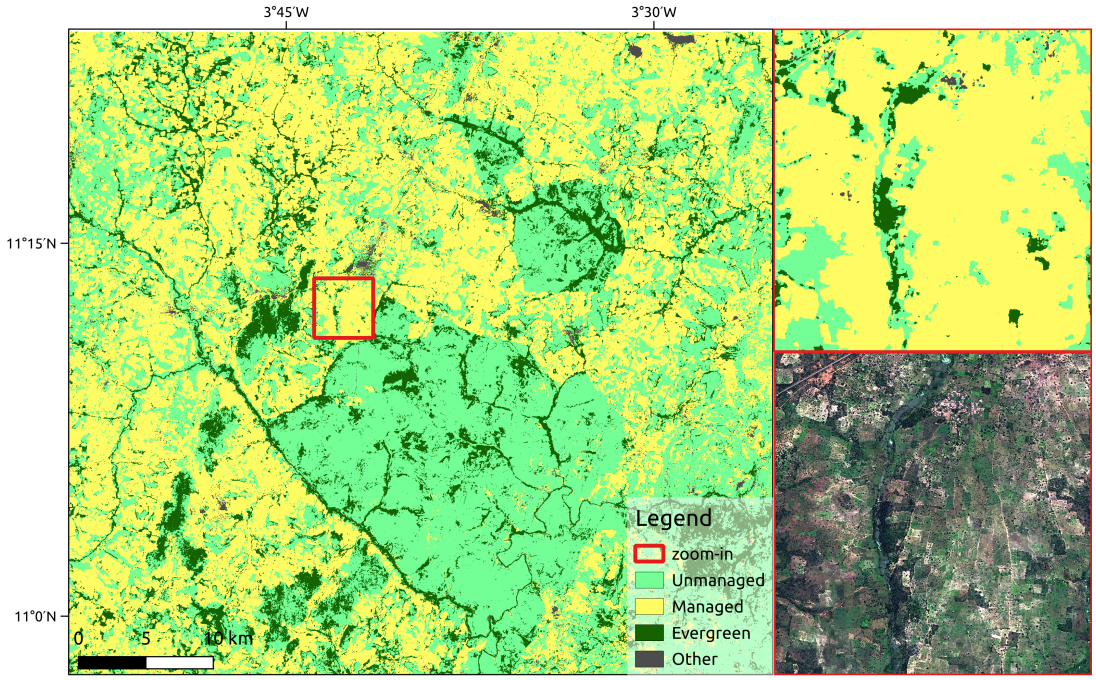


Figure 5.5: Resulting annual 4-class land use map for Koumbia site in 2017. On the right, zoom-in and corresponding SPOT 6 ©Airbus DS 2017. (a) Obtained from RF trial and (b) TCNN trial.

and TCNN where the later did not use rainfall as input data, probably thanks to the ability of the latter to compensate the absence of external sources through the more subtle exploitation of temporal correlations.

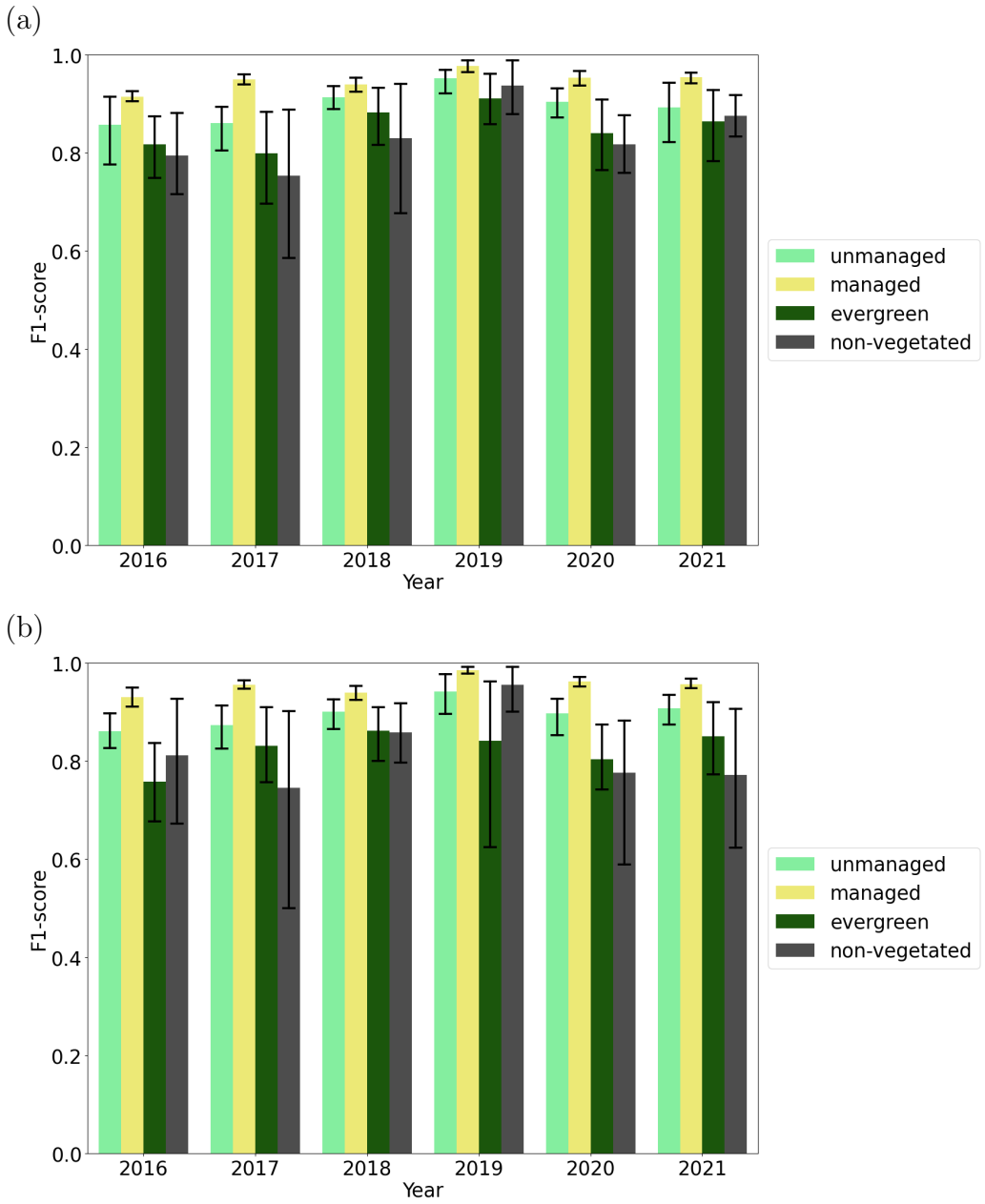


Figure 5.6: Average and standard-deviation (error bars) F1-score of annual management land use class, calculated for the five trained Random Forest (a) and TempCNN (b) model's predictions.

Some interesting elements also come out watching at the maximum class probability map in **Fig. 5.8**, which has also been averaged over years. In**Fig. 5.8** we present model maximum class probability as the average of all folds (5-fold per year, one model per fold), all years confused for the period 2016-2021. For the case of TCNN-based models, in order to highlight potential spatial differences of maximum class probability, values between 0.7-1.0 (top 30%) were normalized. As one might expect, classifier confidence is higher in areas close to the location of reference polygons, even if the average membership probability can also be high in areas where the reference data density is low. However, a pattern of confidence drop is observed across the administrative boundaries

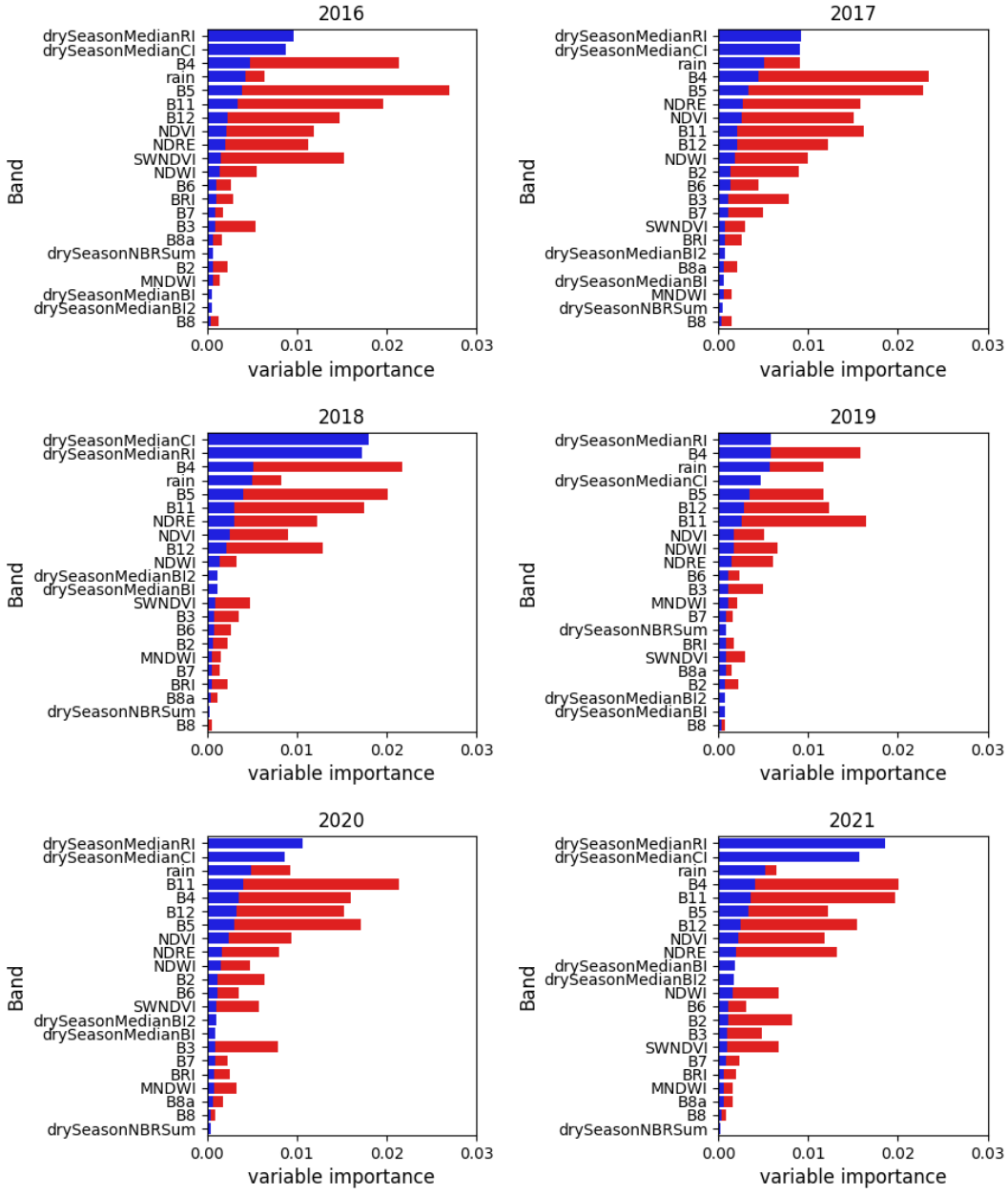
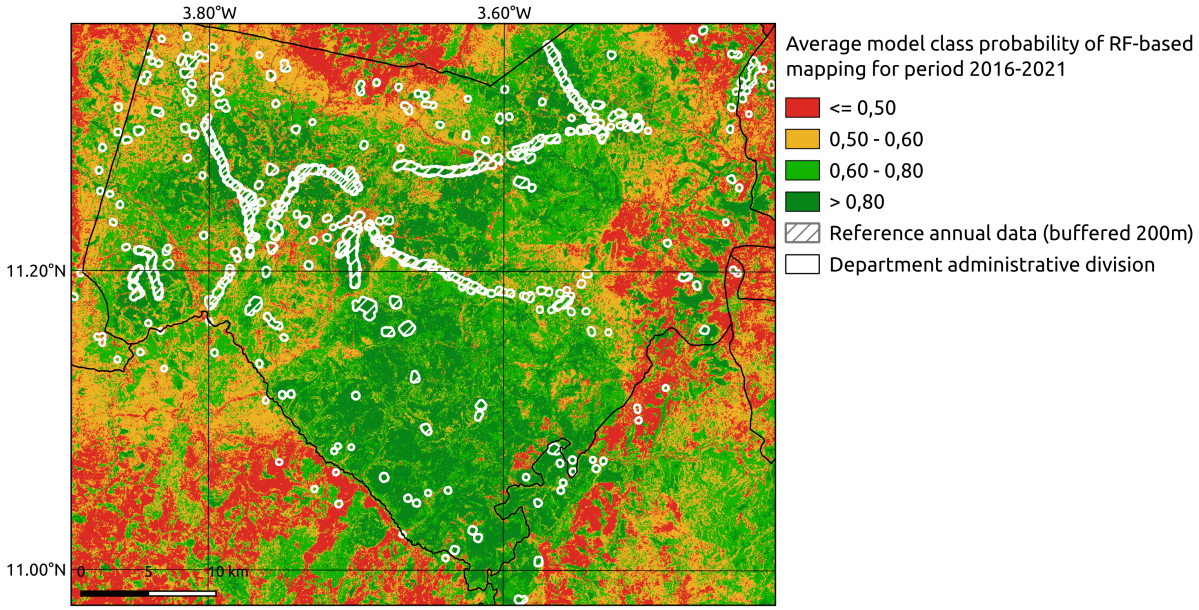


Figure 5.7: Per-year average and importance of variables for the Random Forest land use classification models. The importance values of the 10-day variables averaged over the year are represented in blue, and the annual maximum values are reported in red.

of the Koumbia commune. This spatial pattern may be related to shifts in the landscape outside of the commune, in which reference data loses part of its representativeness. TempCNN tend to yield classification decisions (here used improperly as class "membership probabilities") with much higher levels of confidence (values mainly over 0.8), hence in order to illustrate spatial differences in classification we normalized top values from 0.7 to 1. Thus, TCNN maximum class probabilities present a similar behaviour, resulting in lowest confidence in the south-western quadrant as well as inside the Koumbia commune in areas with a higher density of human settlements, that are characterized by a more heterogeneous landscape.

(a)



(b)

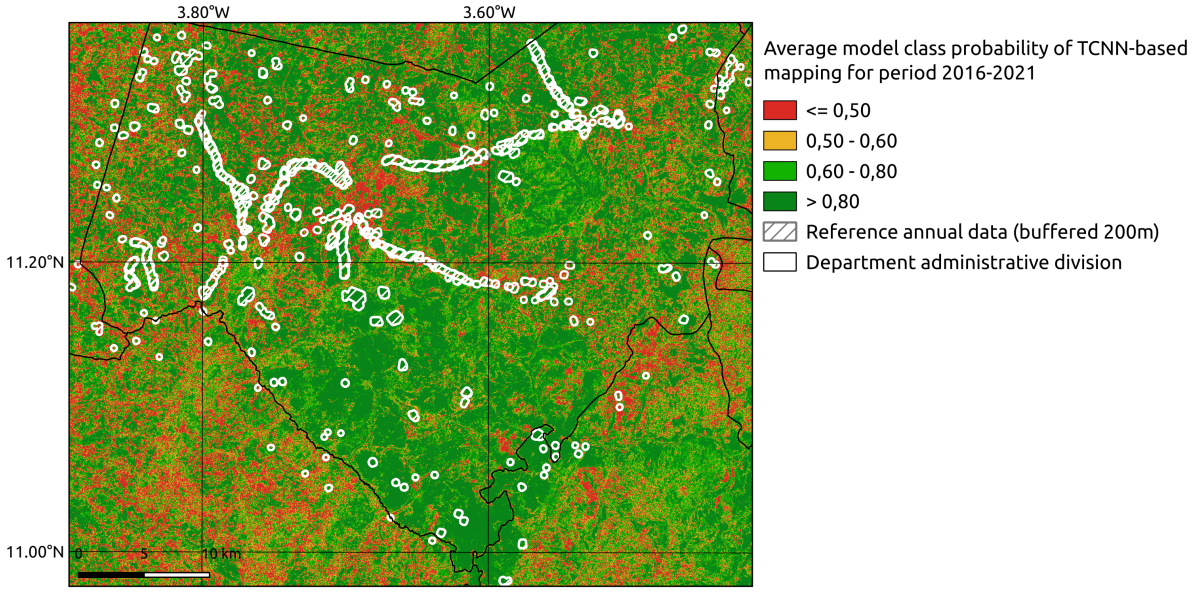


Figure 5.8: Average maximum class probability for (a) Random Forest and (b) TempCNN based mapping for the period 2016-2021 period. Pixel value represents the maximum probability among the four classes considered averaged for all five (5-fold) models and six years considered (2016-2021). For the TCNN case, pixel values represent the top 30% values that have been normalized for enhanced visual inspection. The white polygons correspond to all available years of data from JECAM reference data overlapped with a 200 m buffer.

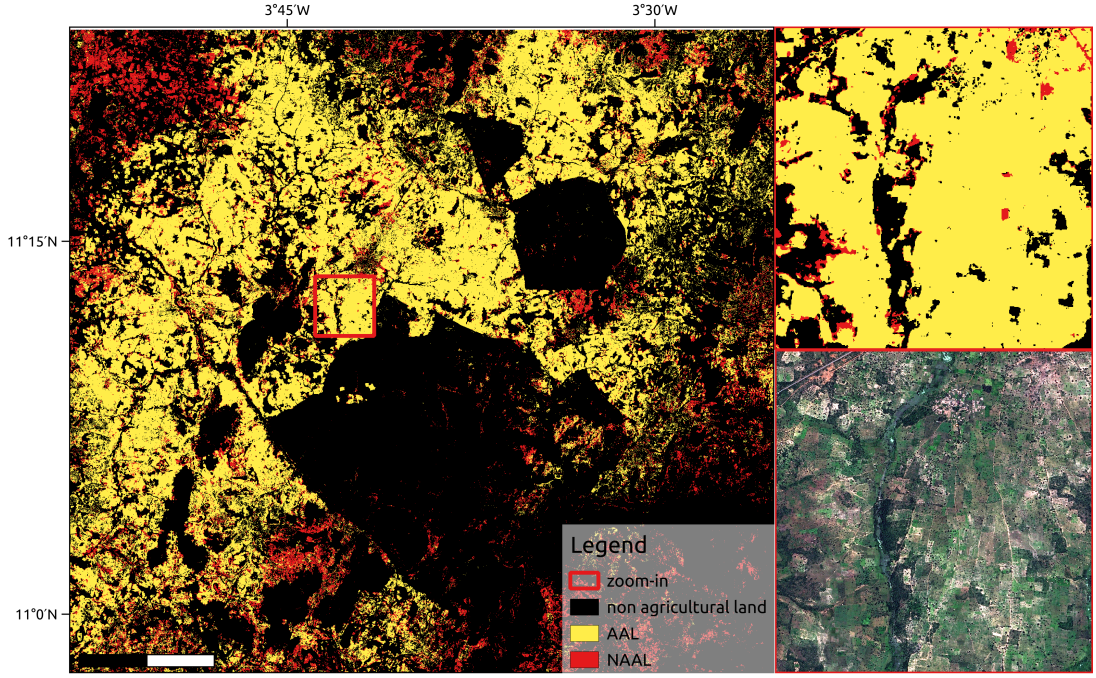
5.3.2 NAAL mapping and validation

Once the 4-class annual maps have been produced for the 2016-2021 time-span, the trajectory analysis described in **Sec. 5.1.2** was applied to provide NAAL maps over the 2017-2021 period. NAAL map for 2017 is depicted in **Fig. 5.9** where, for sake of clarity, the non-NAAL class (O in **Fig. 5.1**) has been split in two classes, respectively the *active agricultural land* (AAL) and the *non-agricultural land*, by simply marking as the latter all pixels that have occurrences of either the *evergreen* or *non-vegetated* classes in the annual land use map time series. Observations made for this year remain valid for other years' NAAL maps as well, which are reported in **Figs. A.1-A.8** (see Annexes), with major instabilities only over the low-confidence area of the annual mapping.

At a first glance, when attending to RF-based NAAL maps, it seems evident that the areas corresponding to low classification confidence for annual land use mapping (see **Fig. 5.9a**) present larger portions detected as NAAL, which are more likely to be due to errors in the base maps than to a true switch in agricultural land use in these areas. Things seem more plausible in the area within the Koumbia commune boundaries, also considering prior knowledge on the study site (NAAL accounts in average for c.a. 10% of the total arable land). However, the zoom in the right-upper part of **Fig. 5.9a** highlights that, apart from some expected salt-and-pepper artifacts, two different spatial contexts emerge concerning NAAL detection, either (a) in the form of suitably shaped fields within AAL (red regular patches surrounded by yellow), or (b) in transition areas between the AAL and the *non-agricultural* class (red irregular patches along black patches). Some details on field-scale NAAL detection for RF trial are depicted in **Fig. 5.10**, where it is possible to appreciate, in different years and spatial contexts, the correspondence between sets of connected pixels (field-like shaped) detected as NAAL in a given year (right column) and the corresponding field appearance on a very high resolution scene (left column) acquired the same year, at the peak of the growing season. In contrast, artifacts on transition areas may be partly caused by errors in the base mapping due to mixed or misaligned pixels in S2 time series. All this makes reasonable the hypothesis of a significant commission error on these results, and motivates the efforts done in further numerical assessment.

If we analyze TCNN-based NAAL maps (see **Fig. 5.9b**) the first major difference is that detected NAAL and non agricultural land (in black) is significantly reduced in those areas where RF trial presented a low confidence, which might reduce commision error for NAAL class. This may indicate that the higher classification confidence of TCNN models was not occasional, but a real sign of a better generalization capability of such technique outside of the reference area. When comparing the zoomed view we can corroborate that (a) many of mapped NAAL in transition areas is (close to to *non-agricultural* land borders) are mainly integrated into *non-agricultural land* or *AAL* class. Moreover, (b) isolated NAAL field-like patches are also present in TCNN NAAL maps. However, a more in depth inspection of this last statement (see **Fig. 5.11**) shows that several NAAL fields initially included (in blue) in JECAM data (and used indirectly for training) are not detected. For the case of fields not included in JECAM data detected in RF trial (in

(a)



(b)

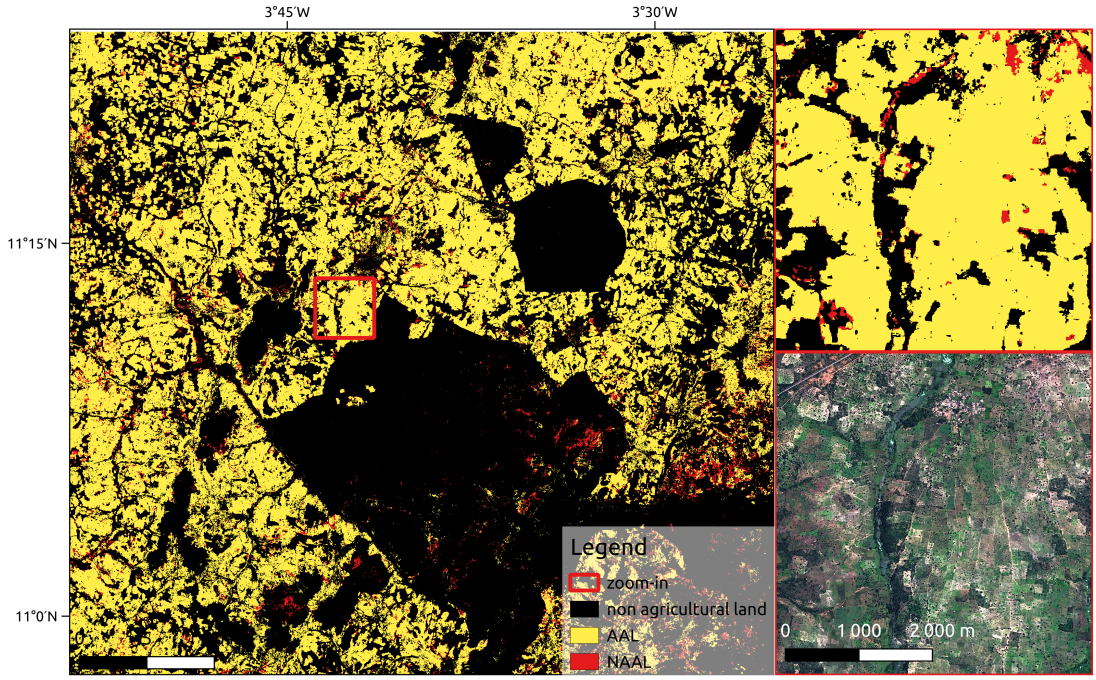


Figure 5.9: Annual non-active land (NAAL) map for Koumbia site in 2017, derived from (a) RF and (b) TCNN trials. On the right, zoom-in and corresponding SPOT 6 image ©Airbus DS 2017.

green), these were either not detected or included into *non-agricultural* land. This last evidence appears to be associated to the previously mentioned transition zones issue and to fields with a high density of trees, which TCNN most likely tend to incorrectly classify

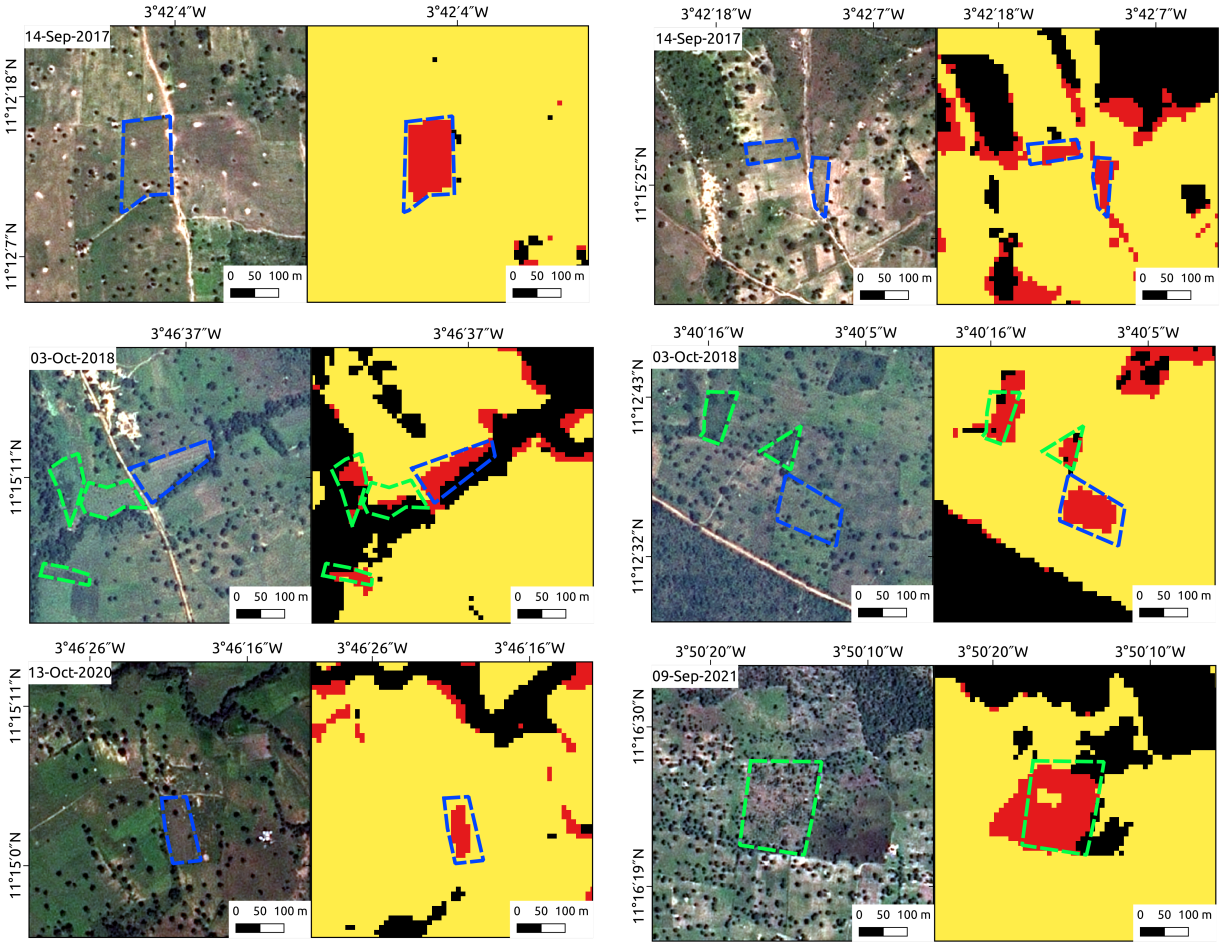


Figure 5.10: Multiple samples of RF-based NAAL mapping at field scale and different years. For each sample, the image on the left is a very high resolution acquisition (SPOT6/7) which has also been used for field delineation in the JECAM data set (Jolivot et al., 2021); Dotted blue lines have been added to highlight photo-interpretated fields for which NAAL reference data was available for that given year. Dotted green lines represents fields for which no reference data was available that given year. All the fields shown here for which a reference polygon existed was correctly labelled as "NAAL".

as *evergreen* class (see example in the bottom right position).

Accuracies (F1-scores) relative to the detection of the NAAL class using the different validation data sets (see **Sec. 5.1.3**) are reported in **Fig. 5.12** for each year (2019 is skipped since no *fallow* samples are available for data set A), and averaged over the period in **Fig. 5.13**, along with uncertainties. A full summary of indicators is reported in **Tab. 5.1**. Note that due to the large class unbalance in all the reference data sets (NAAL only accounts for 5 to 10% of the annotated surface) we only present per-class accuracy metrics for the NAAL class, since the non-NAAL class achieves always a very high score (close to 1).

First of all, non surprisingly, data set A is the one providing the poorest accuracy figures, with a F1-score over NAAL averaged over the period of analysis around 0.53 for RF and 0.58 for TCNN, with the exception of 2017 when a satisfying score of 0.76 for RF and 0.65 for the case of TCNN. One may think that this may be due to the missed

Table 5.1: Annual NAAL maps accuracy metrics (UA, PA and OA stand for user's accuracy, producer's accuracy and overall accuracy respectively) calculated with the five validation data sets.

		RF					TCNN				
Validation	Year[a]	UA	PA	OA	F1-score	Kappa	UA	PA	OA	F1-score	Kappa
A	2017	0.86	0.67	0.99	0.76	0.75	0.79	0.56	0.99	0.66	0.65
	2018	0.44	0.81	0.99	0.57	0.57	0.59	0.77	0.99	0.67	0.66
	2020	0.23	0.72	0.97	0.35	0.34	0.30	0.77	0.98	0.43	0.43
	2021	0.39	0.71	0.99	0.50	0.49	0.47	0.69	0.99	0.56	0.55
B	2017	1.00	0.60	1.00	0.75	0.75	0.99	0.53	1.00	0.69	0.68
	2018	0.99	0.80	1.00	0.89	0.89	0.98	0.79	1.00	0.87	0.86
	2020	1.00	0.61	1.00	0.76	0.76	0.87	0.69	1.00	0.77	0.73
	2021	0.99	0.85	1.00	0.92	0.92	0.96	0.84	1.00	0.90	0.87
C	2017	0.96	0.50	0.98	0.65	0.65	0.95	0.44	0.98	0.60	0.59
	2018	0.97	0.56	0.99	0.71	0.71	0.98	0.54	0.99	0.69	0.69
	2020	0.99	0.57	1.00	0.72	0.72	0.86	0.68	1.00	0.76	0.75
	2021	0.38	0.79	0.99	0.52	0.51	0.48	0.78	0.99	0.59	0.59
D	2017	0.79	0.51	0.98	0.62	0.61	0.75	0.44	0.98	0.55	0.55
	2018	0.65	0.55	0.99	0.59	0.58	0.81	0.49	0.99	0.61	0.60
	2020	0.50	0.79	0.98	0.61	0.60	0.60	0.79	0.98	0.68	0.68
	2021	0.40	0.79	0.99	0.53	0.52	0.48	0.77	0.99	0.60	0.59
E	2017	0.86	0.50	0.98	0.63	0.62	0.79	0.42	0.98	0.55	0.54
	2018	0.68	0.50	0.98	0.58	0.57	0.85	0.45	0.99	0.59	0.58
	2020	0.50	0.77	0.98	0.61	0.60	0.60	0.77	0.98	0.67	0.67
	2021	0.44	0.78	0.99	0.56	0.56	0.53	0.76	0.99	0.63	0.62

[a]Notice that for year 2016 no validation as a result of NAAL detection approach employed (see Sec. 5.1.2)

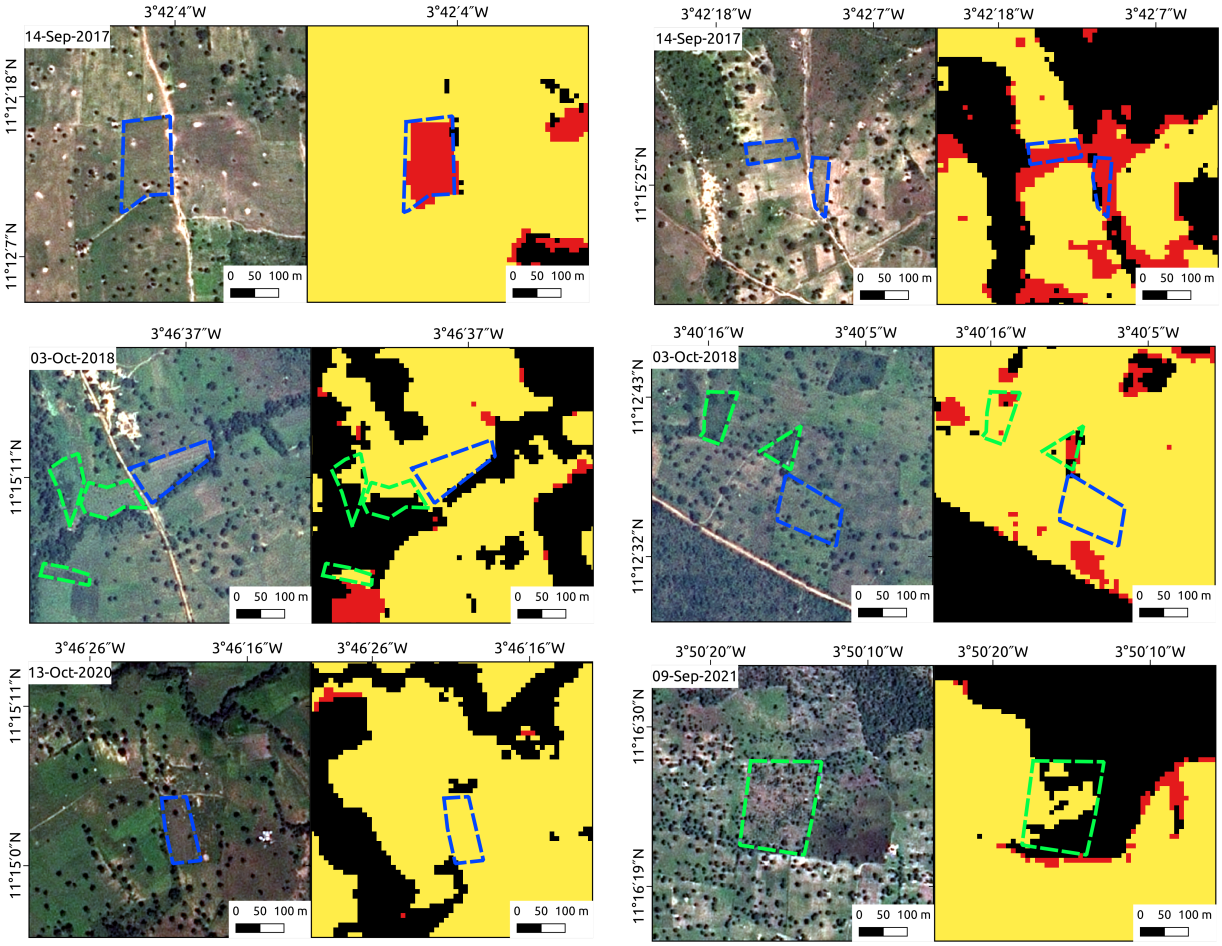


Figure 5.11: Multiple samples of TCNN-based NAAL mapping at field scale and different years. Images correspond to the same fields shown in **Fig. 5.10**. For each sample, the image on the left is a very high resolution acquisition (SPOT6/7) which has also been used for field delineation in the JECAM data set (Jolivot et al., 2021); Dotted blue lines have been added to highlight photo-interpretated fields for which NAAL reference data was available for that given year. Dotted green lines represents fields for which no reference data was available that given year. All the fields shown here for which a reference polygon existed was correctly labelled as "NAAL".

detection (*omission* error) of older fallows, which would in principle require a "bootstrap" period of five years (according to the agreed-upon definition) to be properly covered by our trajectory model. However, looking at **Tab. 5.1** it is clear how the low F1-score is mainly due to a systematically low user's accuracy for both RF and TCNN trials, hence to a high *commission* error. Indeed, in the JECAM database a field is annotated as *fallow* only if a history of active cropping is explicitly known, otherwise non-active cropland fields fall back into the *herbaceous vegetation* class. This again seems to confirm the hypothesis that not all of the NAAL are correctly represented in the original reference database as *fallow* land, making it necessary to validate NAAL mapping with a different strategy. Notice that for the case of TCNN trials, worst results are associated to years 2020 and 2021 suggesting a more accentuated effect of this hypothesis likely due to the lack of *fallow* land data for year 2019 in original reference data set.

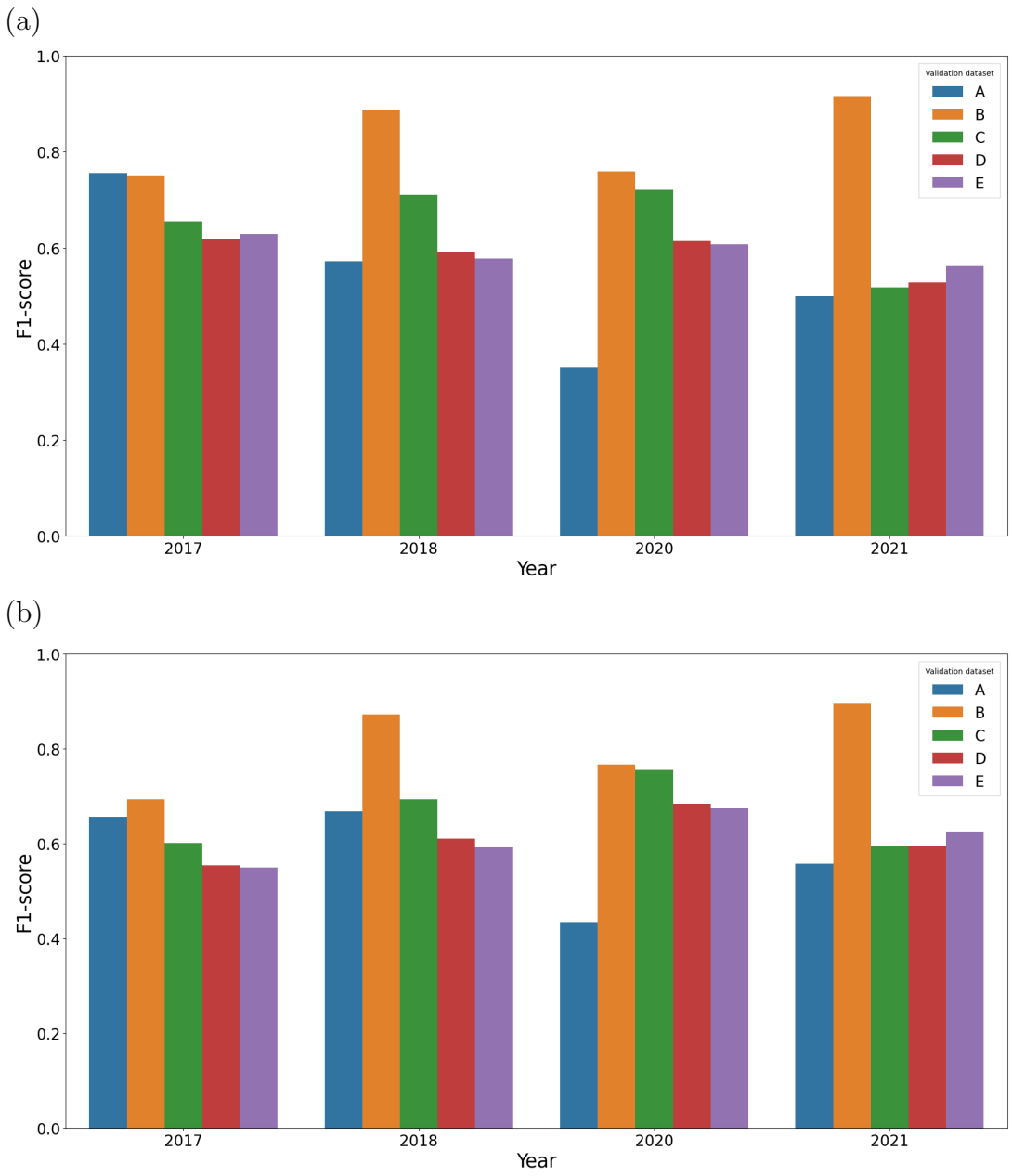


Figure 5.12: Sensitivity of the NAAL class accuracy to the validation data set used (A, B, C, D or E) ; F1-score for RF-based (a) and (b) TCNN-based average annual NAAL calculated over the 2017-2021 period (2019 excluded, see text).

Accuracy improve significantly when validating using data set B (strictly reliable surfaces), with an average score of more than 0.8 over the period for both RF and TCNN trials. However, looking at the number of available reference data points in **Fig. 5.4**, we must face the evidence that this value might not be statistically sufficient, as testified by the larger uncertainty. Moreover, this validation potentially suffers from a positive bias due to the fact that, although the two data sets have a different nomenclature (*managed/unmanaged* vs. *NAAL/non-NAAL*) and data distribution among classes, the validation data points belong to the training surface used for annual land use classification. As expected, accuracy values progressively drop when validating over the augmented vali-

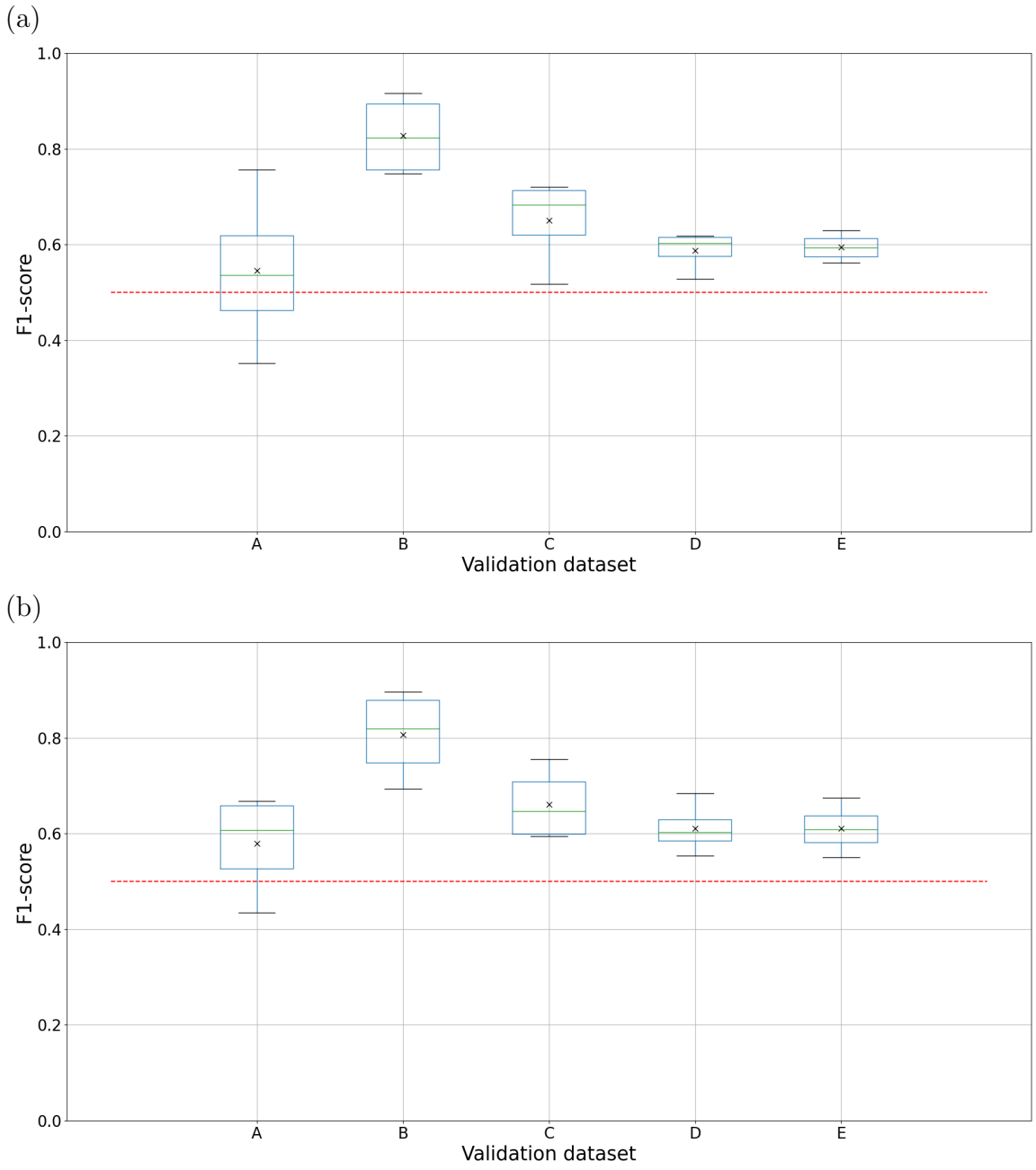


Figure 5.13: Sensitivity of the NAAL class accuracy to the validation data set used (A, B, C, D or E); Box plot of the mean F1-score calculated over the 2017-2021 period for Random Forest (a) and TempCNN (b) trials (2019 excluded, see text). Green line and black cross indicates group median and mean values respectively.

dation data sets C/D/E, attaining average F1-scores ranging from 0.68 to 0.59 for RF and from 0.61 to 0.66 for TCNN trials, but since they rely on a greater number of validation samples (**Fig. 5.4**) the uncertainty is reduced as well.

5.4 Conclusion

In this chapter we explored a trajectory based *NAAL* mapping strategy, where a simplified 4-class based management nomenclature is used for generating annual management maps. Through a restrictive rule system, we discriminate *NAAL* as a change between actively managed vegetated land to unmanaged vegetated land. A thorough evaluation of results was conducted employing different variations of the JECAM data set for the period 2016-2021 as reference data, each accounting for a different assumption made for dealing with a lack of spatio-temporally stable reference data. Accuracy varied depending on the classifier employed, as well as on the validation data set employed with mean F1-scores of over 0.8 using the most restrictive validation data set to 0.6 for when using the most extrapolated validation data set. Geo-spatial quality of predicted maps allowed for identification of "field-like" objects and showed low levels of classification uncertainty at least within Koumbia administrative limits. Significant differences were identified in the classification uncertainty and the level of *NAAL* overestimation depending on classification algorithm employed.

Chapter 6

Multi-year mapping strategies

In this chapter we will test two different strategies for training image classification models using multiple consecutive years of Sentinel-2 imagery (an its associated spectral indices) to identify non-active agricultural land.

6.1 Multi-year approaches

As it has been pointed out along this document, *fallow* practice involves one or multiple years of cropland inactivity, during which natural vegetation regrow and, at a given time, soil fertility might recover to its pre-human activity state. As such, the radiometric features describing *fallow* land or, in a more broad definition, NAAL, might benefit of an enhanced explanatory power when considering temporal periods beyond a single season.

In this chapter we propose two different strategies where we train our image classifier with multiple years of data in order to let the model learn from multi-year dynamics (if any) instead of imposing a set of expertise-based rules. Due to its proven ability of managing temporal correlations and the evidences of a greater generalization capability over the study area, we choose to only test the TempCNN classifier to mine into the information contained in multiple years of Sentinel-2 SITS data. Matter of facts, a Random Forest Classifier basically establish only hierarchical relationships among the input variables, with each variable's discriminative power (i.e., a date of the time series) is tested independently. In other words, just to make an example, NDVI computed from a given acquisition is not explicitly related with any other NDVI, even on closer acquisitions. If this may not be of great limitation for annual time series, whose density is relatively low at the annual scale and whose scattered nature is favored by the presence of clouds, exploiting correlations in time becomes a fundamental prerequisite when analysing multi-year signals. These signals exhibit similar dynamics and seasonality (i.e., periodicity), with the explicit objective of detecting "low frequency" shifts among the different periods, which holds significance at the multi-year scale (e.g., the gain or loss of vegetation productivity through comparison of season-wise cumulative NDVI).

6.1.1 Causal and anti-causal fallow mapping approaches

This approach is related to the one described in the *Direct mapping strategy* (see **Chapter 4**), mainly due to the fact that the employed nomenclature is the same. In other words, we here target the *fallow* land as one of the LULC classes for each given year and simply seek to characterize it through a multi-year satellite image time series. Consequently, a key point of this strategy is the definition of the period of reference to include as input training data, in order to exploit the potential information available to discriminate fallows while (i) keep complexity limited by using only a given number of years and (ii) separately assess the importance of the past and future data with respect to the time series over the target year. This basically accounts to focusing either on the fallow *drivers*, such as the loss of vegetation productivity which may push the farmers to abandon the field or engage a fallowing process, or on the *effects* of fallowing, since the restoration of soil fertility may be measured by satellite data as an increase in the cumulative values of vegetation indices over the next seasons. In all cases, since the objective is to map fallow land on yearly basis, we employ the original JECAM's field spatial delimitation for each year, regardless of possible shifts prior or after the target year in LULC class or boundary limits.

More in detail, we here implement a (a) *causal* temporal approach, where the target year is described by its current and past data, an (b) *anti-causal* approach, where target class is determined by its current and future (with respect to target year) spectral information. Lastly, we also consider a third case (c) which is a combination of previous two approaches, meaning that the LULC class could be described by several years of data centered over the target year, under the hypothesis that both past and future data are likely to be relevant for the task.

6.1.2 Land/Cropping system mapping approach

With the rather straightforward multi-year strategy introduced in the previous section, we are basically seeking for multi-year descriptors that may help the identification of the annual fallow/NAAL class. Again, a major limiting factor may still reside in the lack of sufficient ground truth data for such classes, especially if considering the fact that with larger descriptors (i.e. a bigger number of variables) the need for annotated ground truth theoretically increases (it's the well known *curse of dimensionality*, see Bellman (1966)).

One important factor to consider when mapping any type of NAAL is that it is associated with long-term behaviours that are ultimately reflected into its spectral dynamics at a temporal scale which is larger than a single year. A trivial example may be the difference between two fields, one of which undergoes active cropping, using the same crop types, for several consecutive years, while the other is cropped with cycles of different crop types combined with alternate periods of *fallow*. This suggests a different way to approach the mapping problem that breaks the constraint of the yearly detection fallow land or NAAL, and focuses more to the identification of those portions of cropland on which fallowing occurs, with any frequency, in a given (multi-year) lapse of time. In projection, this may

also be extended to the discrimination of crop surfaces that are managed with a more dynamic approach, with regular changes of practices, in contrast with land undergoing a more speculative approach, to finally assess the performances of these different strategies in terms of productivity in the long run.Nonetheless, such a mapping approach requires an alternative modelling of the classification problem, in which the LULC classes are replaced by a nomenclature made up of different land systems, including *cropping systems* intended as those where cropping practices are implemented over time. Among them, at least one of the classes may represents the system in which yearly NAAL occurs, providing an estimate of the presence of fallow practices at a coarser temporal scale, but possibly more reliable.

Again, based on the JECAM data we defined a novel nomenclature based on land use trajectories, conceived to be associated to spatially stable, multi-year reference surfaces, in which instead of a *fallow* class we target cropland fields exposed (at least once) to *fallow* practice along the period considered. In contrast to this class, we decide to differentiate fields that are always cropped during the entire period. Including areas outside the cropland, we come up with a final 6-class based nomenclature composed of five classes which may be considered stable in time: *non-vegetated*, *evergreen*, *always cropped*, *water* and *herbaceous savanna*, along with a sixth class composed of all those instances where non-active periods emerge after or between two active cropping periods, which we call *dynamic agricultural land* class (DAL). This reduces the cropping system mapping problem to only two classes (*always cropped* and *DAL*), which may be insufficient for a fully fledged characterization of cropping practices over the area, but still may offer an insight to the potential of this approach. Note also that, with specific reference to the aforementioned paucity of reference data on NAAL, this strategy also allows to cumulate reference surfaces where NAAL occurs at least once over multiple years in a single class, hence resorting to a more balanced reference data set as we will show in the next section.

6.2 Implementation

6.2.1 Causal and anti-causal fallow mapping

For these experiments we employ the same 6-class nomenclature used in **Chapter 4**: *cropped*, *fallow*, *water body*, *evergreen*, *herbaceous savanna* and *non-vegetated surfaces*. *Non-vegetated surfaces* is composed of JECAM's *built-up* and *bare soil* land use classes, whereas *evergreen* is composed of *forest* and *savanna with trees* classes.

For both the *causal* and *anti-causal* strategies we use every possible combination of available years in the Sentinel-2 time series as input (see **Fig 6.1**). As an example, when targeting year 2019 using the *causal* approach, we test all the three possible cases which include Sentinel-2 data for year 2019 plus one extra consecutive past year (e.g. 2018/2019 for the first test, 2017/2018/2019 for the second, and so on). For the case of the combined approach we limited the experiment to sets of three years which include

current, past and succeeding year of available Sentinel-2. Finally, as a complementary experiment, we conducted a series of model training tasks where we include all available years of Sentinel-2 data. A scheme summarizing the different experiments proposed in this approach is illustrated in **Fig. 6.1**. For each training job, we make use of the general 5-fold cross validation training protocol as described in **Chapter 2**. Thus, the total amount of data points available for each class and year is the same as shown previously in **Fig. 4.1** (**Chapter 4**), where *fallow* class is significantly under-represented compared to the rest of the LCLU classes considered.

6.2.2 Land/Cropping system mapping implementation

For this experiment we processed all available years of JECAM's data at pixel level (rasterizing data at a 10-meter resolution using Sentinel-2 as reference image) for the period 2016-2021 and applying a trajectory relabelling to infer the nomenclature described in **Sec. 6.1.2** in a pixel-wise manner. For the case of the *DAL* class we included any data point that combined a cropped period with any other vegetated LCLU JECAM class. Thus, if a given pixel has been cultivated at least once and combines either *fallow* or *herbaceous savanna* JECAM LCLU class then it is considered as *DAL*, again assuming possible mislabelling issues in the source JECAM data. *Non-vegetated* and *water* classes are not allowed to be mixed with other vegetated classes. Hence, when these two classes are present in the period considered, *water* or *non-vegetated* class is preserved. A scheme of the implementation of this nomenclature is shown in **Fig. 6.2**.

Subsequently, we performed a regrouping of connected pixels with the same class labels and proceed to split data to perform a 5-fold cross validation strategy as described in **Chapter 2**, with the difference that data splits are done using these newly generated objects. Distribution of data points per class is shown in **Fig. 6.3**. The total amount of available data points using this nomenclature results into a more balanced distribution, at least between the critical *always cropped* and *DAL* classes, which are by the way significantly less represented compared to *herbaceous savannas*.

6.3 Results

6.3.1 Causal and Anti-causal fallow mapping approach

We conducted a series of trainings using TempCNN classifier and a 5-fold cross validation approach as described in **Chapter 2**. Thus, we trained an ensemble of five models for every possible combination of Sentinel-2 annual data for causal (C), anti-causal (A), the combination of causal and anti-causal (CA) and the complementary trials for which all available Sentinel-2 data is used (ALL). This yielded a total of 31 training jobs for all possible combinations of years (not including year 2019 since no *fallow* data is available in reference data), all experiments combined. Average 5-fold F1-score for *fallow* class is

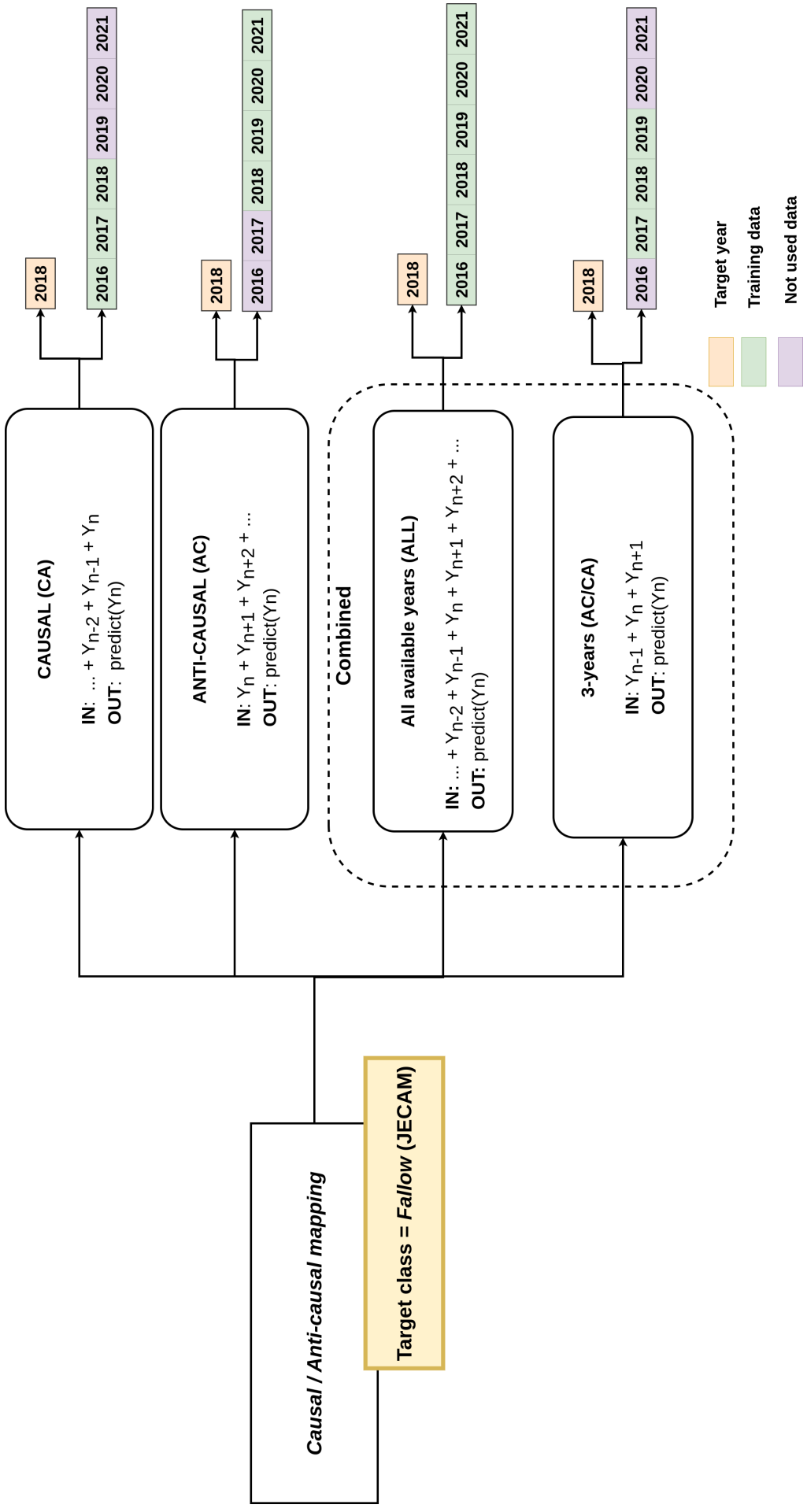


Figure 6.1: Scheme of all training strategies applying a *Causal/Anti-causal* approach.

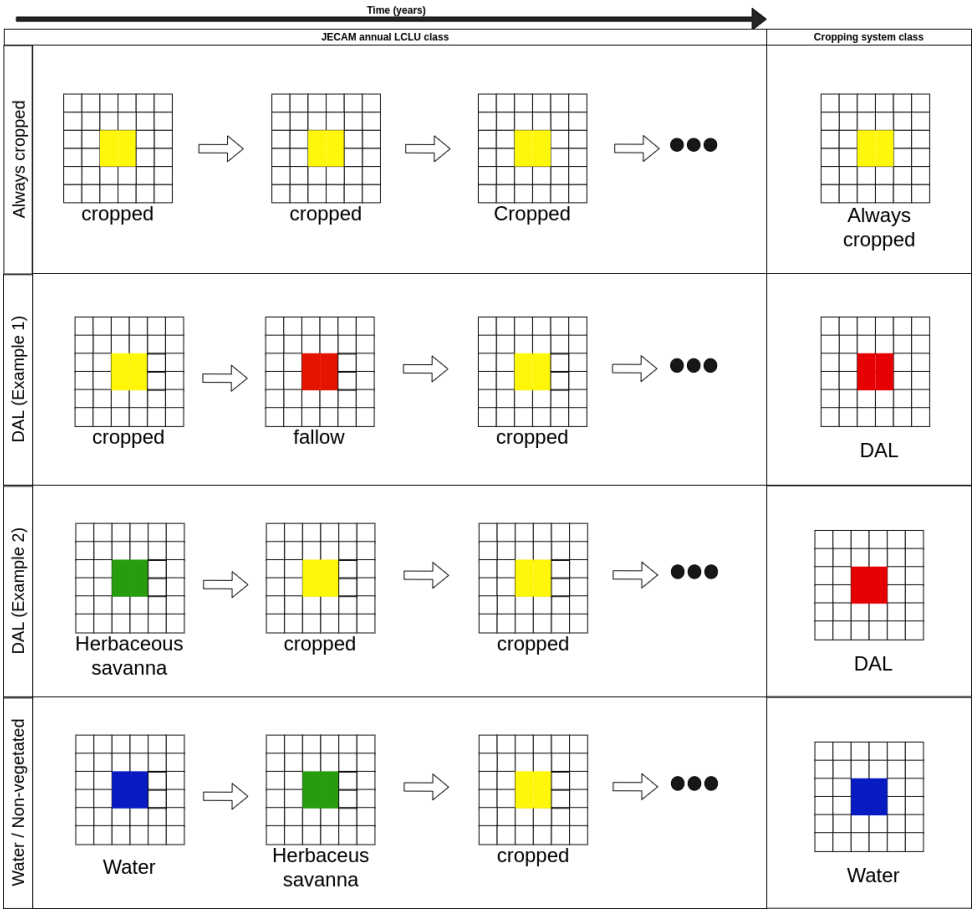


Figure 6.2: Scheme of application of *Land/Cropping system* nomenclature (*always cropped*, *dynamic agricultural land*, *water*, *evergreen*, *non-vegetated*, *herbaceous savanna*) at pixel level. Both *non-vegetated* and *water* classes apply the same overwriting principle when present in the trajectory.

shown in **Fig. 6.4**, where F1-scores reached a maximum value of 0.42 for year 2020 in trial $C3$ (i.e. trained with 2020, 2019 and 2018 data).

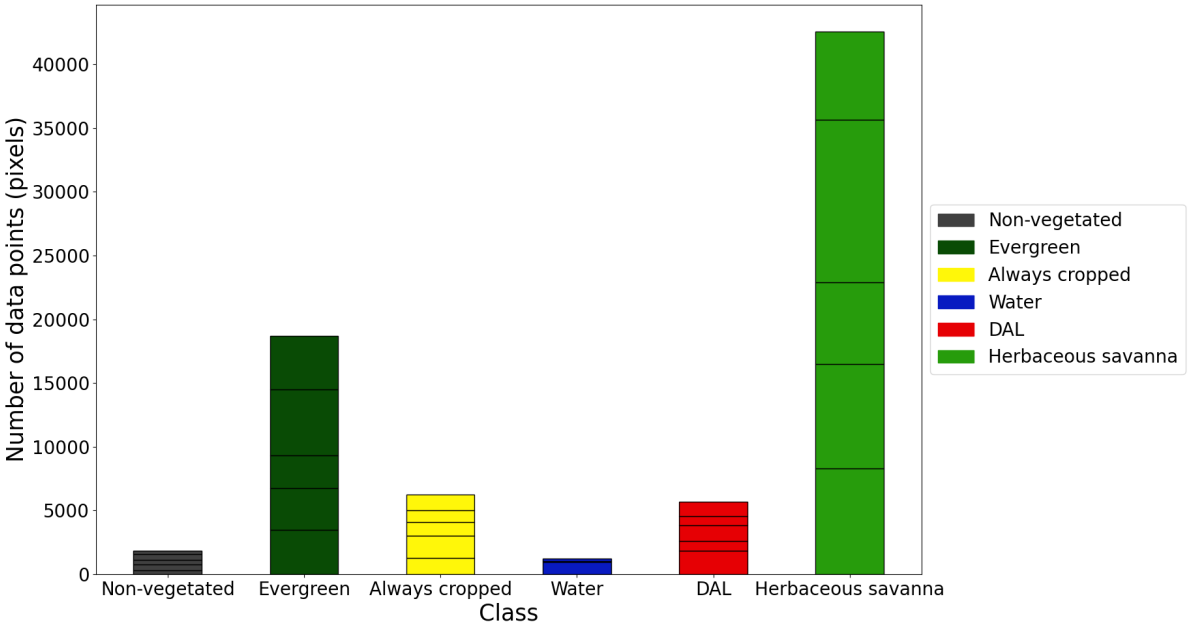


Figure 6.3: Number of data points in pixels (after rasterization using the Sentinel-2 10-m grid as spatial reference) per class. Bar splits represent the proportion of pixels employed in each cross validation fold.

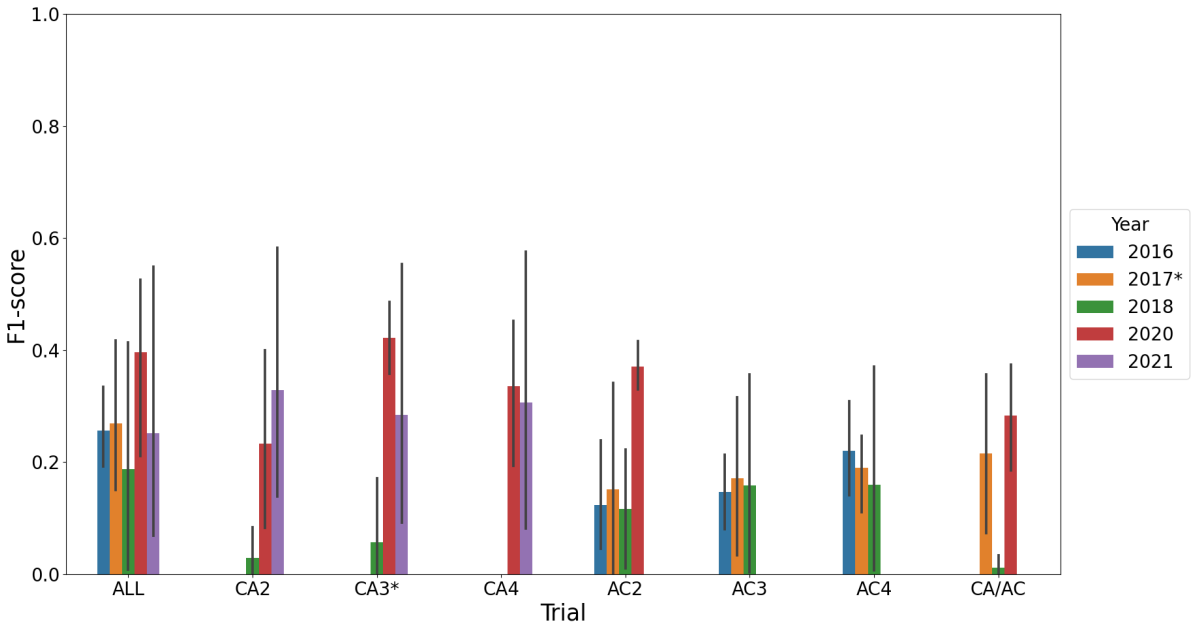


Figure 6.4: Average 5-fold *fallow* class F1-scores for causal (CA) and anti-causal (AC) multi-year approaches. Number coding shows how many years of Sentinel-2 data were included in training (ex: $CA2$ includes two consecutive years of data). ALL indicates that all available years were included, whereas CA/AC is composed of sets of 3 years (Y_{n-1}, Y_n and Y_{n+1}). Black error bars represent f1-score standard deviation. *Trial $CA3$ for 2017 is not shown because all folds yielded an F1-score of zero.

Results presented in **Fig.6.4** show that independently of the strategy (causal/anti-causal) or the amount of data used for model training ($C2$, $C3$, ...) targeting directly *fallow* land does not allow for sufficiently accurate results, with performances with high

variations for the same target year with multiple instances were F1-score is below 0.1. Average F1-scores for each training strategy was 0.27, 0.22, 0.17 and 0.23 for *ALL*, *CA*, *AC* and for the three year combined causal/anti-causal *AC* respectively. In **Fig. 6.5** we present the classified map of one of these trials as an illustrative example.

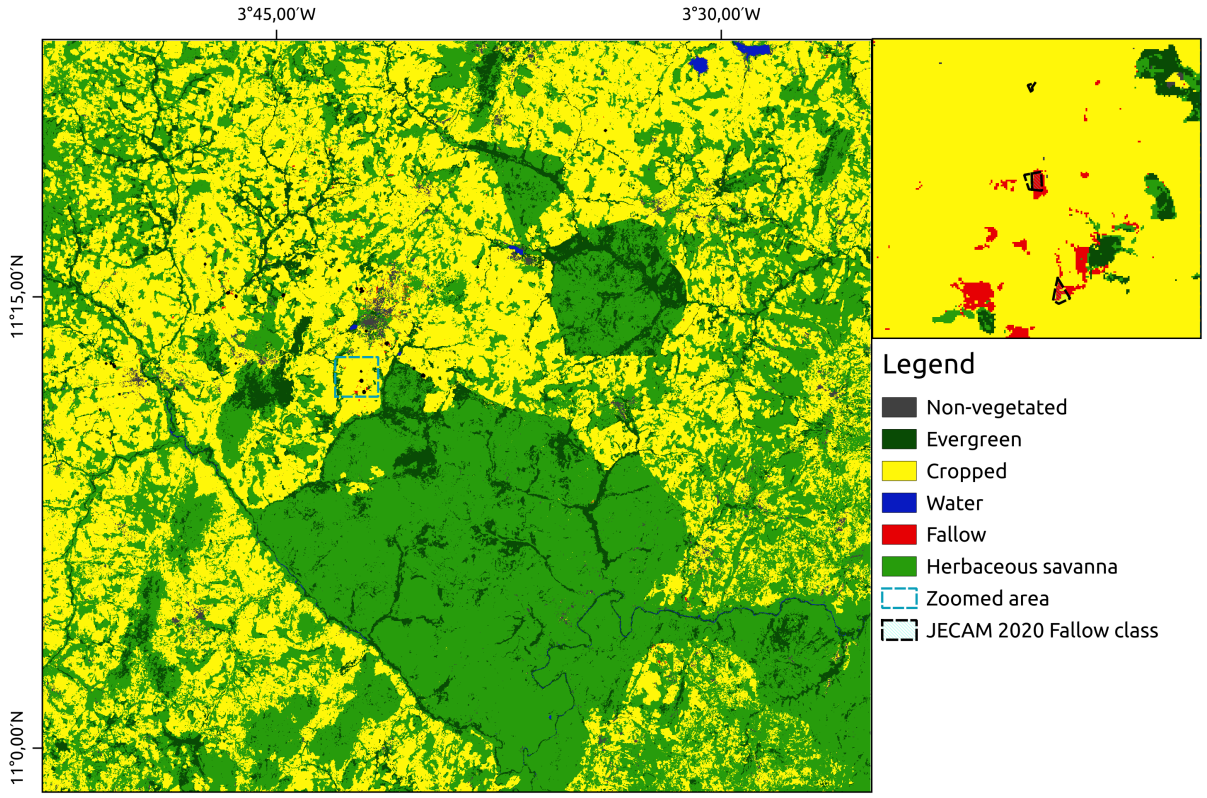


Figure 6.5: Example of predicted map (5-fold average) for 3-year *causal* trial (CA3) for year 2020.

In the example of **Fig. 6.5** we can quickly notice that the *fallow* class is rather absent, with known *fallow* fields as registered in JECAM data set mislabelled as *cropped* (see top right zoom). The spatial distribution of *cropped*, *evergreen* and *Herbaceous savanna* appears to be coherent to other image classification trials in previous chapters. In fact, if we attend to the per-class confusion matrix shown for this example in **Tab. 6.1**, we detect overall acceptable performances, with F1-scores over 0.75 mark for classes *evergreen*, *cropped*, *water* and *Herbaceous savanna*. Besides *fallow* class, low F1-scores are also detected for class *non-vegetated*. A significant commission error is found in *evergreen* class (30% *CO*), with mislabelling issues occurring between classes *evergreen* and *herbaceous savanna*. For the case of *fallow* class, most of commission and omission errors occur towards *cropped* class. Note also that the two worst performing classes (*non-vegetated* and *fallow*) correspond to those for which the less amount of reference data was available.

Table 6.1: Average 5-fold confusion matrix of *3-year causal* ($CA3$) 2020 trial testing results. *CE* and *OE* stand for *commission error* and *omission error* respectively, calculated over average confusion matrix, whereas *UA*, *PA* and *F1* stand for user accuracy, producer accuracy and f1-score respectively.

Reference / Predicted	Non-vegetated	Forest	Cropped	Water	Fallow	Herb. savanna
Non-vegetated	148.80	0.20	38.80	0.00	10.20	59.60
Forest	0.00	2131.60	16.00	0.00	0.20	951.60
Cropped	35.20	6.80	5712.40	0.00	51.40	64.60
Water	3.60	0.00	0.00	200.00	0.00	39.00
Fallow	3.40	0.20	91.80	0.00	63.40	17.40
Herb. savanna	87.20	508.60	132.80	0.00	7.20	5805.20
avg. CE	0.35	0.27	0.03	0.19	0.64	0.12
avg. OE	0.33	0.19	0.05	0.00	0.45	0.18
avg. UA	0.67	0.81	0.95	1.00	0.55	0.82
avg. PA	0.65	0.73	0.97	0.81	0.36	0.88
avg. F1	0.59	0.76	0.96	0.83	0.42	0.85

If we now consider the all the classes (see **Fig. 6.6**), we can observe that, with the exception of *fallow* class, all other classes reach F1-scores beyond the 0.7 mark. For the case of *cropped*, *herbaceous* and *water* there are no significant differences among the trials, with one exception for *water* class in trial *CA3*. Despite the fact that it remains within acceptable ranges (lower whisk above 0.8 mark), we can reasonably state that these classes do not benefit from adding extra input data beyond the limits of a single season. When we shift our focus to classes *evergreen* and *non-vegetated*, we observe a contrasted and opposite behaviour between them, where average F1-scores for *causal* trials are higher than *anti-causal* trials in *evergreen* class, whereas the inverse relationship is observe in *non-vegetated* class. For the case of *fallow* class, we no longer observe clear differences between *causal* and *anti-causal* trials. Nonetheless, we observe a cumulative increase of median F1-score when incorporating extra years of input data for both approaches (CA and AC). Notice that highest F1-scores for *fallow* class were obtained with the *causal* approach, which also yields the highest variability. Nonetheless, despite the relative enhanced performance for *fallow* class, the absolute F1-scores remain sub-optimal with respect to the land trajectory strategy introduced in **Chapter 5**.

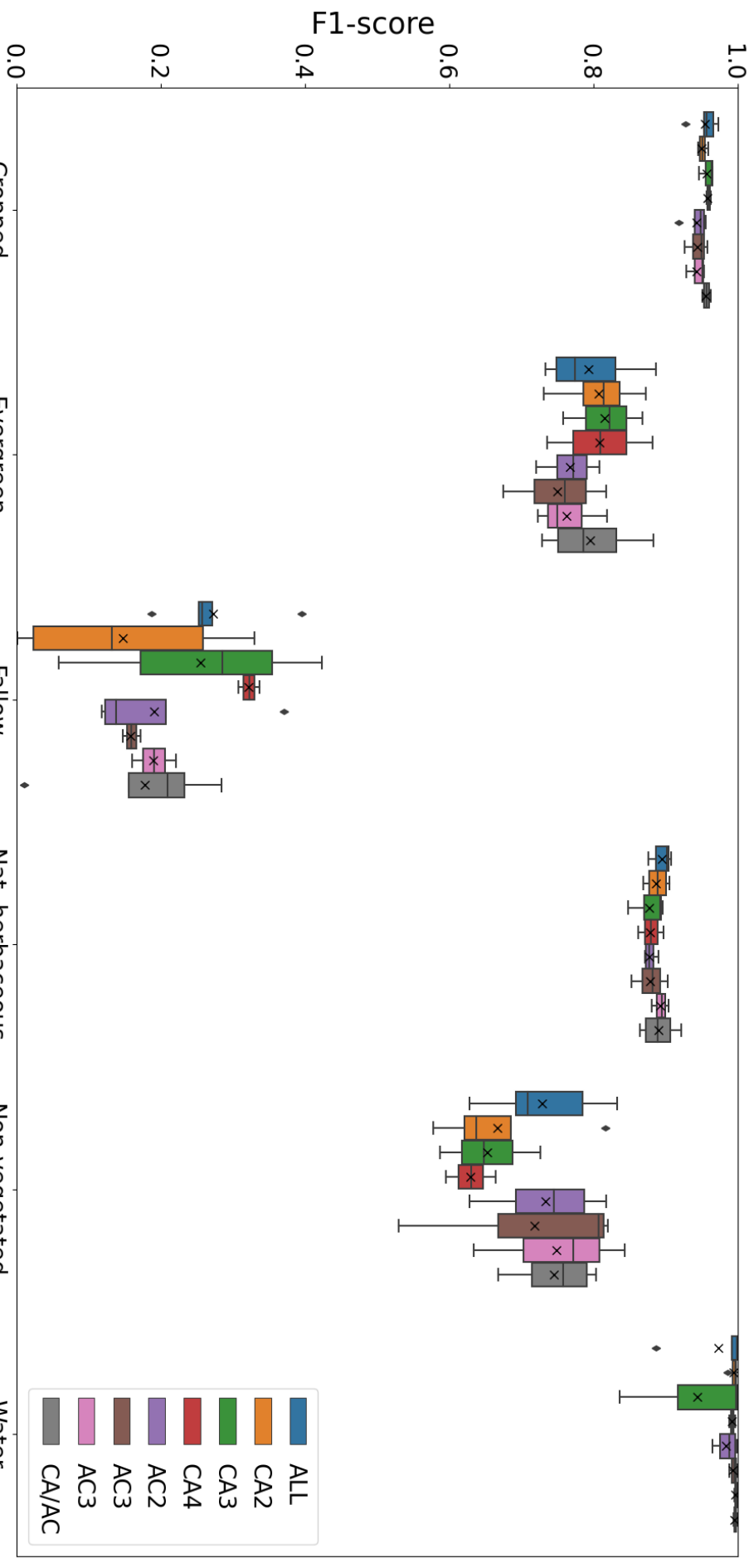


Figure 6.6: Average 5-fold F1-scores for all classes by approach causal (CA) and anti-causal (AC) multi-year approaches. Number coding shows how many years of Sentinel-2 data were included in training (ex: $CA2$ includes two consecutive years of data). *ALL* indicates that all available years were included, whereas CA/AC is composed of sets of 3 years (Y_{n-1}, Y_n and Y_{n+1}).

6.3.2 Land/Cropping system classification

We conducted a 5-fold cross-validation following the strategy described in **Sec. 6.1.2**, using all available years of Sentinel-2 data and its indices for the period 2016-2021 to train a TempCNN model. Resulting model obtained an 5-fold average overall accuracy (OA) of 77.69% and per class F1-scores ranging from 0.44 to 0.86 (See **Fig. 6.7**). Target *DAL* class obtained the lowest average F1-score of 0.44.

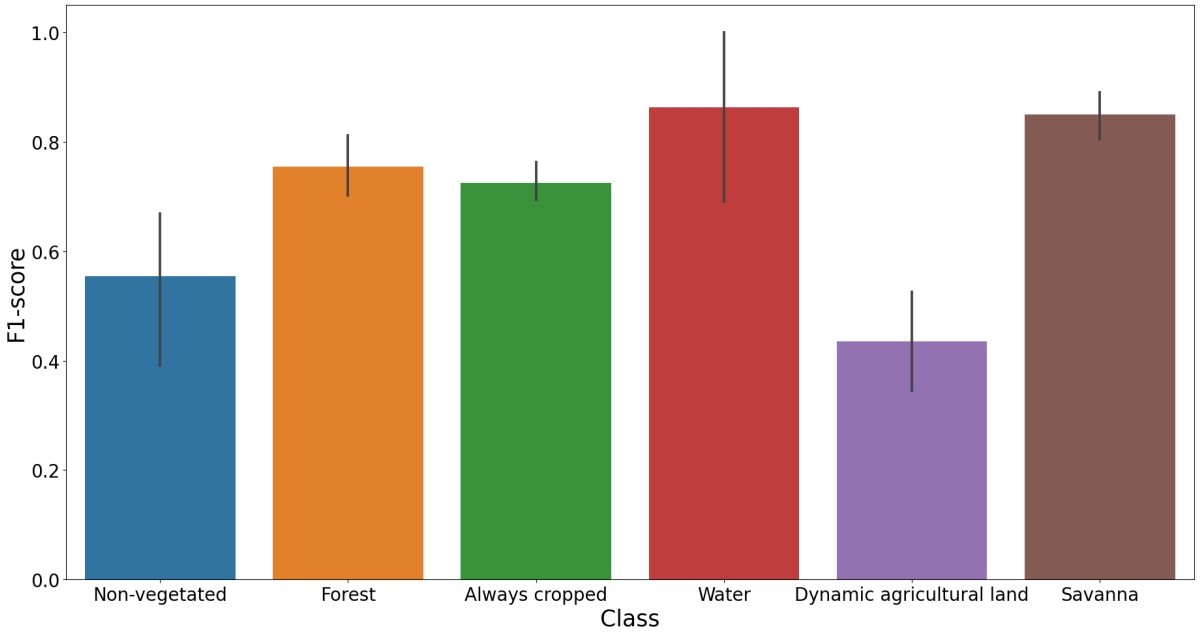


Figure 6.7: Average 5-fold F1-scores for multi-year Land/Cropping system mapping strategy.

Clearly, the proposed strategy proves ineffective even considering the presence of fallows cumulatively over the Sentinel-2 coverage period. Anyway, despite the sub-optimal results obtained for this mapping strategy, this approach did yield improved performances when compared to previous causal/anti-causal strategy where, although the classes are not directly comparable, *fallow* land obtained average F1-scores below 0.3 independently of the training approach used. Taking a look at the average confusion matrix (see **Tab. 6.2**), we can see that both types of errors (commission and omission) are present and most conflicts are related to the *always cropped* and *savanna* classes, which is clearly due to the proximity in spectral dynamics among these three classes, which may be enhanced by the fact that, in given periods of time, they may account for basically the same condition in terms of vegetation growth (crop activity or inactivity).

In **Fig. 6.8** we present an example of the classification results obtained using this mapping approach. At first glance, and as expected, there is high presence of *DAL* with a proportion over total cropland (i.e. *always cropped* plus *DAL*) of 75%. It is worth recalling that, due to the nomenclature change this values are not directly comparable to JECAM annual *fallow* land class. Nonetheless, we can perform a coarse estimate over

Table 6.2: Average 5-fold confusion matrix of *Cropping systems* testing results. *CE* and *OE* stand for *commission error* and *omission error* respectively, calculated over average confusion matrix, whereas *UA*, *PA* and *F1* stand for user accuracy, producer accuracy and f1-score respectively.

Reference/Predicted	Non-vegetated	Forest	Always cropped	Water	DAL	Herb. savanna
Non-vegetated	183.40	12.00	8.20	0.00	**67.80**	95.60
Forest	0.20	2965.80	2.80	0.00	**146.00**	625.00
Always cropped	6.00	3.20	887.00	0.00	**316.20**	38.60
Water	0.00	0.80	0.00	239.60	**1.20**	1.00
DAL	**24.60**	**79.60**	**256.80**	**0.00**	**563.20**	**210.80**
Herb. savanna	55.40	1056.20	49.80	0.00	**335.00**	7016.20
CE	0.50	0.21	0.29	0.01	**0.50**	0.18
OE	0.32	0.28	0.26	0.00	**0.61**	0.12
UA	0.50	0.79	0.71	0.99	**0.50**	0.82
PA	0.68	0.72	0.74	1.00	**0.39**	0.88
F1	0.58	0.75	0.72	0.99	**0.44**	0.85

JECAM data set by stacking all ever cropped and fallowed surfaces for the period 2016-2021, which yields a ratio "at least once fallowed" over cropland of 3.1%.

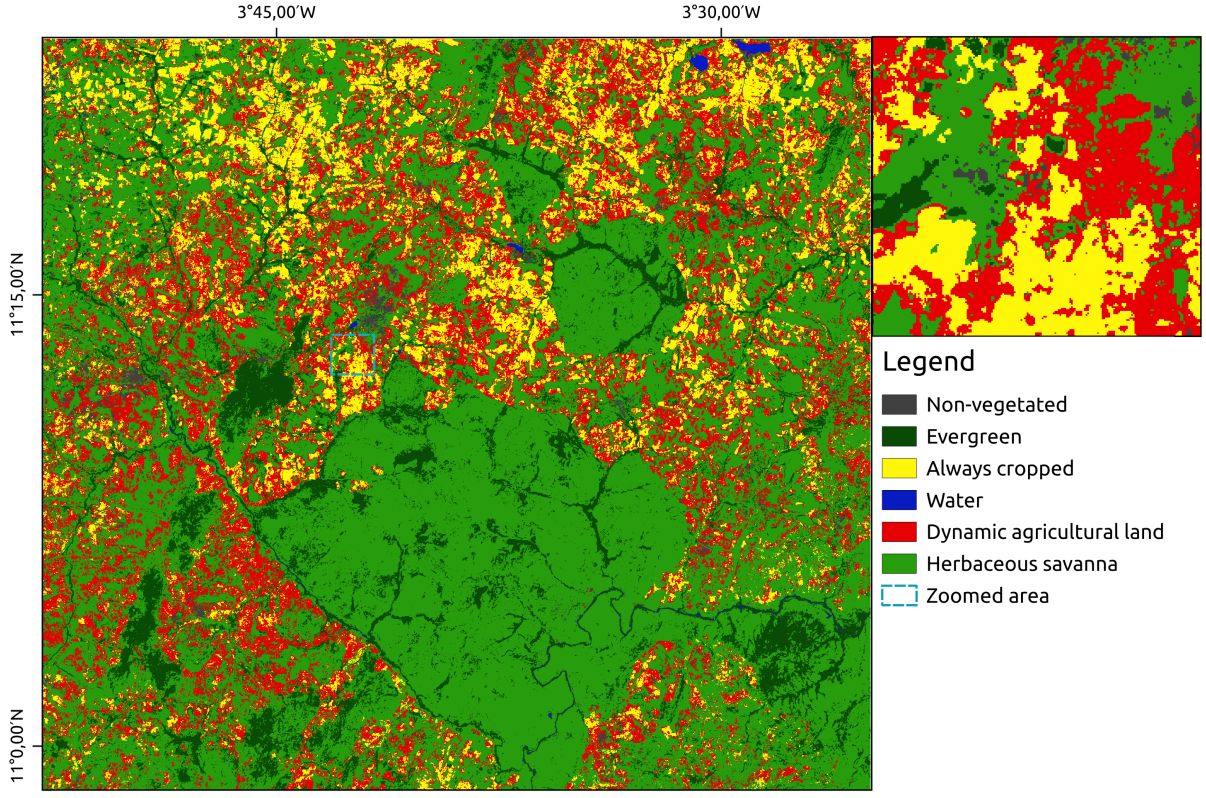


Figure 6.8: Predicted classification results for *Cropping system* approach.

6.4 Conclusion

On this chapter we tested two different approaches for training a deep learning TempCNN model in which we employed multiple years of Sentinel-2 as training data. In the first causal/anti-causal approach we target *fallow* class as included in JECAM database, which yield poor performances independently on the type or quantity of data included during training. For the second approach, accounting for land/cropping system mapping, we used a more wide nomenclature and targeted what we called *Dynamic Agricultural Land* (DAL), which is not linked to a given year/season but encompasses the whole period of analysis. This last approach yielded better results than the first multi-year approach, reaching the 0.44 mark on average F1-score for *DAL* class, whereas the majority of multi-year *fallow* mapping trials did not reach the 0.3 F1-score mark. Although the cropping system approach yielded marginal improved performances for the target fallow-related class, none of the approaches guarantee a sufficient level of accuracy as for providing reliable estimates of *fallow* land.

Chapter 7

Discussion and conclusion

7.1 The path from land cover to land use

7.1.1 Challenges and issues

The work presented in the different chapters composing this document aimed at proposing different strategies for the automatic mapping, using time series of optical multi-spectral satellite images, of *fallow land*. The targeted land is basically made up of land surfaces which undergo what is at its core, an agricultural decision (i.e. not cultivating that portion of potentially cultivated land), whether it be direct or indirect. As pointed out progressively across the work, such research challenge presents a series of inexorable constraints, from the necessity of detecting a given practice that is not properly defined by any particular observable state of land (i.e. land cover, LC), to the additional difficulties associated to our specific agro-climatic context.

Starting from the latter, the agro-climatic characteristics of the study site are indeed the first contextual element that set the amplitude of our challenge : with rainfed agriculture and precipitations concentrated in a few months, vegetation growth both for cropland and naturally vegetated spaces happens in the very same lapse of time, creating an unfavorable conditions for satellite imagery based analysis with an overlapping of spectral signatures and a reduced amount of cloud-free images. Moreover, the reduced size of fields (often less than 1 ha), a strong heterogeneity in the management of cropped surfaces in both intra- and inter-annual strategies (heterogeneity of cropping practices, reshaping of fields, etc.), the presence of trees and weeds in cropping fields as well as the variations in soil fertility, are all factors that limit the potential of remote sensing imagery in characterizing *fallow* land.

Additionally, this highly dynamic agricultural system is also being affected by an increasing population growth, which requires an increase of cultivated land to subsist, but rather causes a decreased availability in financial and human resources for the agricultural sector, creating situations of land abandonment as a consequence of the lack of inputs to manage it. In fact, this means that yet another kind of land dynamics could be related, although inappropriately if considering a more strict definition, to fallow practices. The

lack of resources is also revealed in more indirect manners, with few paved roads enabling the transport and movement of goods and workforce, or the absence of a technological data infrastructure storing official graphical registers of land use (and hence potential high quality reference data on larger scales). These are just some of the factors that describe a highly fragmented and dynamic landscape that represents a wide portion of Sub-Saharan Africa, and that might help the reader fixing the ideas on the complexity of the mapping problem and understanding the exploratory approach that we adopted in this thesis .

7.1.2 An exploratory approach

In this thesis, our exploratory approach intended to leverage the use of multiple years of radiometric data obtained from time series of satellite images, eventually complemented with geo-spatialized rainfall (CHIRPS, Funk et al. (2014)), to derive discriminating information for the mapping *fallow* practices by means of different single- and multi-year image classification strategies, elaborated by coupling machine and deep learning techniques to suitable knowledge-based modellings of the classification problem. To this aim, we first made the essential choice to make use of Sentinel-2 imagery as a well adapted product that can capture seasonal vegetation dynamics and allow us to work at the field level, with specific reference to the tropical/sub-humid West African agrosystems in which our study area is located. We also decided to make use of the JECAM data set, since it provided us with multiple series of ground truth data that could be employed for (i) characterize the presence of fallow on the study area and the potential drivers that could be observed to identify fallow practices, and (ii) build suitable reference data sets for training and testing the machine learning building blocks supporting all our mapping strategies.

In **Chapter 3** the limitations mentioned in the previous paragraph were made explicit by analyzing JECAM's *fallow* samples in combination with Sentinel-2 derived time series of vegetation indices (NDVI), showing that this type of land results indistinguishable from other types of agricultural land uses when exploiting data at an annual scale. Enlarging the time-span of analysis and working at field level permitted to detect instances where NDVI appeared to show a positive inter-annual NDVI increase when *fallow* practice was being used. Yet, these patterns were not general enough on the whole data set to allow for a possible rule-based system for *fallow* land detection. This lack of generality is probably due to the highly heterogeneous fields, where it is possible to have vegetative states and dynamics that vary significantly from one field to another, as a results of pedoclimatic conditions, but also as a consequence of extended periods of seeding/harvesting and the absence of any kind of mechanization, without mentioning the sparsity of available satellite observation during the growing season. The potential of data on soil types as a possible predictor of the occurrence of fallows has also been tested, unfortunately with poor outcomes.

As a consequence of this acquired awareness, we decided to set the remainder of the work around the main research question on: How can we transition from land cover

information, defined as the present state of land with properties that are measurable through satellite imagery, to a land use understood as an abstract categorization of land described by its exploitation purposes, as exemplified by fallows in their agricultural system? Due to the lack of any evidence on the existence of *direct spatial indicator* for the presence of fallow, we opted for a more data-driven strategy based on the use of machine learning (mainly Random Forest and TempCNN classifiers), and tested different strategies to model *fallow* land.

As a first *baseline* approach, in **Chapter 4**, we tested different direct annual mapping methods, where fallows are addressed as another land cover class. In this phase, our reference surfaces are those annotated as *fallow* in the JECAM data set, although well aware of the noise that they may contain (missed - mostly - and wrong fallow annotations). Results showed that in all instances, targeting *fallow* land directly as a machine learning/deep learning classification problem is not sufficient to differentiate this type of land use, with average F1-score below the 0.2 mark. Annual strategy results exhibit high levels of omission for the *fallow* class using both RF and TCNN classifiers. For sake of completeness, we also tested an unsupervised classification strategy recently appeared in literature (Tong et al. (2020)) specifically conceived for fallow mapping in the semi-arid (Sahelian) band using Sentinel-2 imagery, which as expected performed poorly in our area (which is located in the subhumid band), resulting in significant *fallow* class overestimation (commission error). This goes in line with previous trials of direct annual mapping of *fallow* land on the same area, even if in a larger land cover mapping scenario, in which the *fallow* class was mostly missed (average F1-scores lower than 0.2) (Gaetano et al., 2019).

It soon appeared clear that employing annual time series of radiometric features does not provide sufficient information for a classification model to "learn" to identify the *fallow* practice, due to the absence of a specific signal for the related land cover dynamics. Truth is, this is not surprising given the multi-year nature of this practice, whose occurrence depends not only on the observed degradation of land productivity after multiple years of active cropping, but also on multiple pedoclimatic and socioeconomic factors which end up widening the range of *fallow* land types, finally making their yearly characterization practically impossible. Moreover, the strategy described in **Chapter 6**, namely the *causal / anti-causal* approach, which aims at detecting fallow land in each given year through the use of multiple subsequent years of Sentinel-2 acquisitions, also proved rather inaccurate. Taking all necessary precautions into account, it is highly likely that this evidence suggests that the long-term degradation of vegetation growth, potentially leading to fallows, as well as its regenerative effects, are not consistently captured by the input time series. This renders the "naive" multi-year approach ineffective. Therefore, we esteemed necessary to adapt our strategies to cope with this variability and exploit the most identifiable properties of this practice.

We hence decided to revisit the findings of our exploratory study on the reference surfaces, which further confirmed the hypothesis that vegetation dynamics on fallow land

develop over extended periods rather than within a single year. This motivated us to adopt an alternative approach, one that transcends the modeling of *fallow* practice solely as a distinct type of land cover. Instead, we interpret it more coherently as a *land use* type, necessitating a more explicit definition based on observable characteristics through satellite data. This rationalization brought us to the broader concepts of *Non-Active Agricultural Land* (NAAL) and *Dynamic Agricultural Land* (DAL) employed in **Chapter 5** and **6**, that respectively represent to two ways to "extrapolate" from proper fallows in (i) a land use class which includes all surfaces which are not cropped in a given year but have been cultivated in the recent past (*de facto* including abandoned fields) and (ii) an intrinsically multi-year land use class encompassing all surfaces where, in several subsequent years, active cropping is alternated to idle periods (hence mapping a *cropping system* that makes sense only over multiple years). Note that, the solution (i) still accounts for a yearly mapping of NAAL, but the constraint of "rigorous fallow" is relaxed, while in (ii) we clearly conceptualized a different mapping product, whose reference period might be of several consecutive years, focusing mainly on the identification of agricultural practices which are susceptible to make (direct or indirect) use of fallows to regenerate land productivity.

Summarizing, in order to develop a somewhat viable approach for detecting *fallow* land we conceived different manners to conceptualize it, rethinking what are the characteristics that define a particular use of land and how it can be isolated. We concentrated the problem around the temporal dimension, which yielded different conceptualization (see **Fig. 7.1**) from more "traditional" definitions of *fallow* as a land cover class, to more "atemporal" definitions where *fallow* practice is encompassed within a highly variable agricultural system. Following rigorously constructed validation protocols, which utilize classical machine learning tools and appropriately modify the original JECAM reference dataset, transitioning from annual approaches to multi-year modeling and strategies proves to be highly beneficial for mapping performance. Of these strategies, the land trajectory modeling approach (NAAL mapping) stands out as the most effective to date.. However, many are the questions to be raised on the viability and generalization potential of these approaches, which we will try to analyze in detail in the reminder of this chapter.

7.2 Exploiting temporal dimension

7.2.1 Direct multi-year mapping strategy

In **Chapters 4** and **6** we tested different mapping strategies that target annual *fallow* land using JECAM's *fallow* records as reference surfaces. This allowed us to established a mapping "baseline" that implements a similar logic as in more traditional land mapping strategies, where targeted classes can be directly explained by a given dynamic of radiometric features along the target year (often the growing season). As for the *causal/anti-causal* approach this was extended beyond a single growing season and tested different

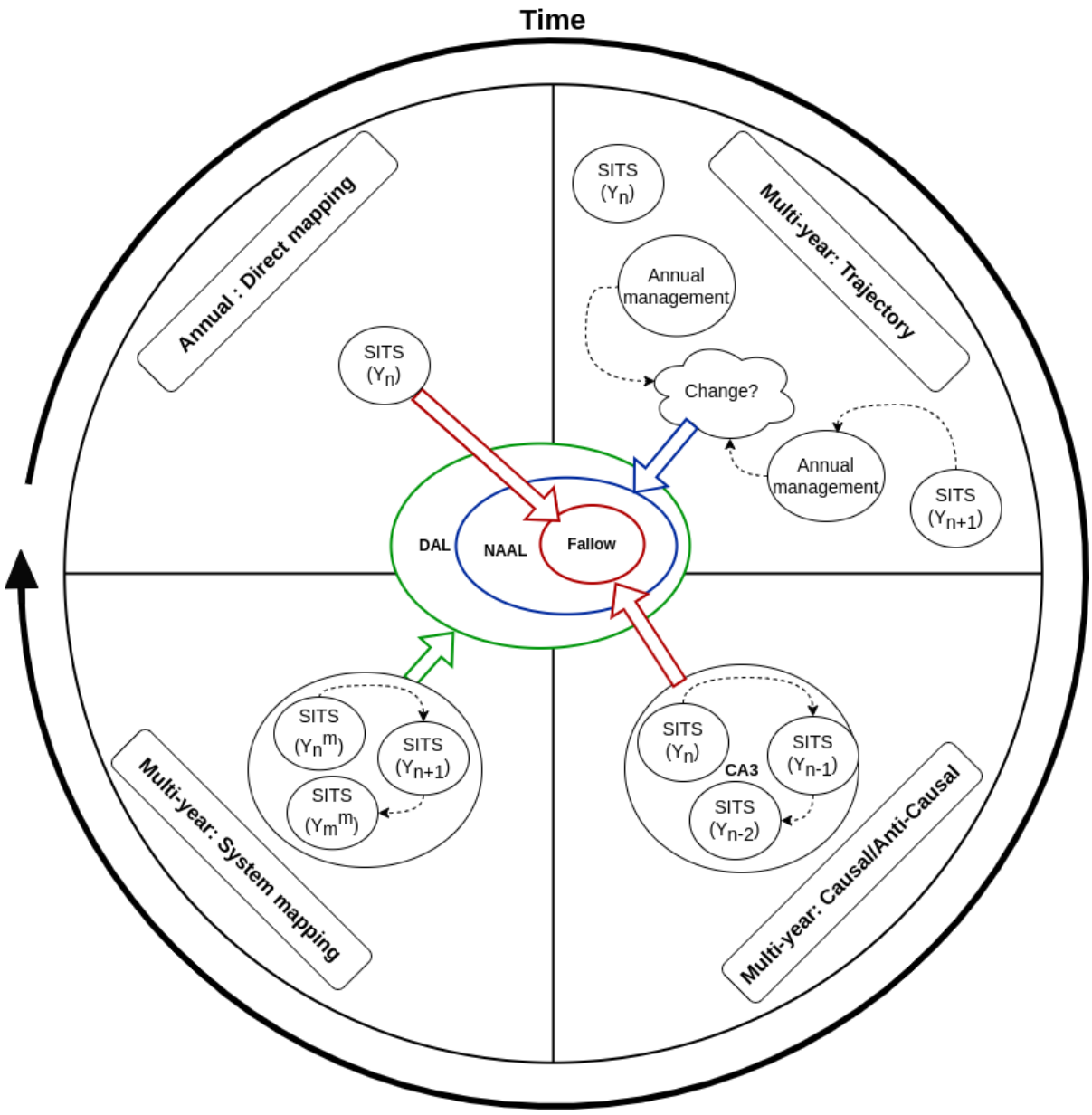


Figure 7.1: Summary scheme of all the approaches tested and objective target. *NAAL*, *DAL* and *SITS* refer to *Non-Active Agricultural Land*, *Dynamic Agricultural Land* and *Satellite Imagery Time-Series*. Each *SITS* represent the available images for a given year (or agricultural season), where n is the current year and m the total amount of *SITS* available.

types of temporal relation w.r.t. the target year. In **Chapter 4**, classification trials were not capable of accurately detect annual *fallow* land. Nonetheless, other classes considered achieved satisfactory classification accuracies with F1-scores beyond the 0.6 mark, showing that for some types of land use direct approaches as these might suffice.

In **Chapter 6** the *causal/anti-causal* approaches have shown the effects of using extended periods of input data, in the past and future directions and both, on the very same classification workflow. Results were contrasting for the *fallow*, *evergreen* and *non-vegetated* classes (see **Fig. 6.6**), with surprisingly the latter class present enhanced F1-scores for *anti-causal* (*AC* trials). This is probably related to the fact that classes with a more stable radiometric dynamic over the years (this one includes JECAM's *built-up* and *bare soil* classes) may benefit more from the redundancy of the input signals (since it raises classifier's confidence). Although this may also be the case for the *evergreen* class (exhibiting similar temporal profiles every year), this effect was more unstable, providing significantly improved F1-scores only for the *causal* trials w.r.t. *anti-causal* trials, possibly indicating a certain sensitivity in classes that may be affected by shifts in seasonal rainfall. For the most important case of *fallow* land, there were no clear differences between *causal* and *anti-causal* trials, yet we did observe slightly enhanced F1-scores when extending the time period considered of the input radiometric features, suggesting an increased explanatory power with longer time-series.

7.2.2 Trajectory analysis based strategy

As mentioned above, results showed in **Chapter 5**, might be the most viable way to provide reliable mapping of NAAL surfaces based on Sentinel-2 image time series in the sub-humid West African agrosystems. An extensive validation has been provided using multiple versions of the reference data set, showing that if we keep the most reliable information available on NAAL reference surfaces the measured accuracy is very high (> 0.8 F1-score) considering the fine scale targeted. Moreover, when coping with possible biases in such data set, results also show that, enabling different amounts of reference data augmentation based on suitable hypothesis on spatial correlation at field level, performances still remain significantly moderately positive (F1-score ranging from 0.5 to 0.8) if compared with results achieved in direct mapping strategies (**Chapter 4** and *Causal/Anti-causal* approach in **Chapter 6**. Due to the lack of other reference databases long enough to further validate this methodology, this assessment does not yet allow an operational transfer of the proposed method as a decision support tool, nonetheless we may legitimately state that trajectory-based model proposed here achieve promising results.

Spatial inspection of NAAL maps also confirms that (i) many areas detected as NAAL are compatible to agricultural fields in shape and size (see **Fig. 5.10**) and (ii) the overall occurrence of NAAL surfaces, which amounts in average to around 11% of the total arable land (with peaks of 14% years 2019 and 2021) is coherent with the occurrence

of fallow fields in the JECAM data set across the period. Thus, we can conclude that the proposed methodology is adept at generating locally significant NAAL maps down to the plot scale. This stands in contrast to recent studies on fallow mapping (Tong et al., 2020; Dara et al., 2018; Rufin et al., 2022), which primarily concentrate on large-scale assessments and do not offer evidence of comparable local significance at the level of pixel aggregates. Yet, there is room for improvement in spatial accuracy which could be achieved by post-processing NAAL maps based on prior information on the location and shape of the detected fields.

7.2.3 Cropping systems mapping strategy

The results of the *land/cropping systems mapping* approach (**Chapter 6**) where it is attempted to locate the so-called DAL surfaces provided improved results when compared to annual and multi-year *fallow* land direct mapping. Although the *fallow* class from the original JECAM data set and *Dynamic Agricultural Land* class are semantically different, the latter provides an alternative way to quantify the presence of fallow in a given agricultural system, which is less precise in time but still might give precious information for its assessment. However, overall performances obtained in *cropping systems mapping* strategy were not sufficient for generating a reliable cartography, with an average F1-score of 0.42. In this approach land use change detection is being handled directly by TempCNN algorithm, which is designed to exploit the temporal dimension in a multivariate intra-seasonal/annual time-series. Hence, might not be properly adapted for multi-year land use change detection. In this line, more traditional multi-year analysis tools such as *LandTrendr* or *BFAST* (Verbesselt et al., 2010; Kennedy et al., 2010), oriented to detect seasonal trends and requiring long time series (20+ years), may achieve better performances. Although a possible way to implement such strategy could be the use of longer time series of decametric satellite images (such as *Landsat* imagery at 30 m resolution), we doubt that such data may offer the adequate spatial and temporal resolution to cope with the specificity of our targeted agricultural system.

7.3 The importance of ground truth data

In all the classification tasks conducted in this thesis we used the JECAM reference data for testing the resulting accuracy of generated maps allowing for the assessment over the same study site along the period 2016-2021. This data set, which has the advantage of providing explicit records on *fallow* fields, allows us to develop and evaluate different approaches for mapping *fallow* practice, as well as to identify the limitations, requirements and methodological adaptations needed to this aim. However, it is important to also understand what are the limitations related to the reference data itself: quality and extensive ground truth data sets are rare, resource intensive and logistically challenging, which motives the development of "cheaper" solutions including those who leverage satel-

lite imagery to upscale to larger areas. Yet, the resulting accuracy and quality of these mapping strategies is to be tested against high quality reference data, ideally recorded *in situ*, which makes initiatives for building ground truth data such as JECAM data set indispensable.

Herein we leverage such data set, but also explore its limitations through the different classification tasks performed. Certainly, the availability of a suitable reference database, both precise at field scale and conducted over multiple years, is rather uncommon for the target region. On the other hands, we also could measure that the non-static nature of field boundaries reduces the surface available for the validation of multi-year strategies, making it necessary, as we have shown in **Chapter 3**, to process the original samples in order to build a spatio-temporally harmonised ground truth data. Additionally, the limited accessibility of large portion of the study site also limits the potential coverage of the data set: gaps are present in the multi-year characterisation of a significant percentage of fields, which implied further noise-prone processing to data set prior to its use. But most importantly, the representativity of collected samples, which are almost mostly collected along passable pathways, is questionable. In **Chapter 5**, some evidences are provided on how this fact can directly impact the confidence of classification models, whose reliability drops when moving far from the areas where reference data is available. In other words, ensuring the spatial representativeness of reference data is one extra desirable feature that can enhance model predictive capabilities by capturing spatial differences linked to changes in soil properties, climate (e.g. uneven rainfall), vegetation or other factors. Always concerning representativity, another fundamental limitations to fully assess the proposed strategies relies on the absence of reference data on other study sites, at least within the same geographical area: investing in the construction of several other data set enabling for the multi-year tracking of cropping practices is necessary to provide further evidences on the methodological clues brought up by this thesis.

One may argue that the building of such data sets could be addressed through the use of high resolution imagery and photo-interpretation. Yet the feasibility of this strategy would rely on the ability of discriminating fallows, or at least their encompassing classes (NAAL, DAL), from other types vegetated spaces, including cropland and natural spaces. This condition is rarely met in sub-humid West Africa due to a mainly rainfed-based agriculture and a lack of reliable cropland mapping products (without mentioning the limited availability of exploitable image acquisitions during the rainy season). Any improvement in image spatial resolution or temporal frequency within the current range of available products are therefore unlikely to provide substantial improvements both in the construction of an extensive enough reference database through photo-interpretation or in providing improved radiometric quality with potentially enhanced explanatory power.

7.4 Perspectives, limitations

7.4.1 Multi-sensor approaches to cope with cloudy images

As explained in Chapter 2, in this thesis we have dealt with the presence of clouds in the satellite image time series by means of a linear "gap-filling" technique to infer missing values, thanks to the available cloud masks provided. Yet, this method, which is necessary in order to provide complete time series to the classifiers, performs poorly especially during the rainy season, where a significant number of consecutive missing acquisitions produces a piece-wise linear signal which can only hardly reproduce the real changes happening under the clouds. One potential alternative to alleviate the lack of cloud free imagery is the integration of images from different sensor, including products obtained by the fusion of multiple imagery, e.g. MODIS-Landsat (Gao et al., 2017), Sentinel-LandSat (Wu et al., 2022), where the reduction of cloud presence is achieved by an increased combined image revisit time. Yet, this is still dependent on the occurrence of cloud-free days during the rainy season, which may be rare, without mentioning the implicit reduction in either radiometric quality or pixel spatial resolution that the fusion methodologies imply. More advanced solutions worth considering are approaches where Synthetic Aperture Radar (SAR) images, such as those provided by the open-access Sentinel-1 mission with its cloud-penetrating capabilities, may be used jointly to optical data to improve classification accuracies, or even fused by means of machine and deep learning algorithms in order to generate estimations and "fill" missing radiometric features of interest such as NDVI (see Scarpa et al. (2018); Ebel et al. (2020); dos Santos et al. (2022); Lasko (2022)). An Alternative focus can involve more complex cloud removal strategies ("gap-filling) through the use of deep learning algorithms and radar data (see Meraner et al. (2020); Zhang et al. (2020)). Thus, showing that there is a wide range of state-of-the-art methodological improvements that can potentially enhance image/data quality.

7.4.2 Moving towards unsupervised methods

Matter of fact, in this thesis we provide a first evidence that for the identification, at least a rough one, of NAAL surfaces the regular detection of active cropland vs. other kinds of non-managed deciduous vegetation may suffice, since this would enable the application of a trajectory model like the one proposed in Chapter 5. Hence, a possibility that we consider worth exploring to address the issue of large-scale NAAL identification is to resort to unsupervised classification techniques for the *managed vs. unmanaged* LULC mapping problem, eventually adapted to specific agro- and eco-systems. One way to proceed could be to exploit the proposed supervised classification workflow to deeply inspect the combination of image-derived variables allowing for the discrimination between these two classes, and design a proper unsupervised methodology to provide such discrimination. In this setting, the deployment of field work for gathering reference data may be based on an agrosystem-scale sampling in representative areas, moving the whole workflow within

the range of feasibility.

An intermediate solution towards building an approach less reliant in high quality ground truth data and also increase the extent/area covered is to use strategies relying on *Positive Unlabelled Learning* (Bekker and Davis, 2020), a particular machine learning setting where the learner only knows about a "positive" class and deduces the negative one from the difference in the unlabelled set. In our case, the positive class could be the active cropland, the *unmanaged* surfaces being the negative class. In this scenario we could limit the field efforts to collecting samples of active cropland, generally easier to access and annotate, and let the machine manage the antagonist class prior to proceed to a trajectory based analysis.

7.4.3 Methodological improvements

Finally, in this thesis we employed TempCNN image classifier as an state-of-art deep learning alternative to Random Forest classifier, which showed statistically similar performances in all classification trials tested herein, but yielded improved spatially coherent classifications in areas where no reference data was available (see **Chapter 5**). Moreover, due to the intrinsic potential within the "layered" structure in deep learning models it seems worth exploring more adapted or *ad-hoc* deep learning models where user can, to some extent, "guide" the model and determine what are the building blocks from which to derive the explanatory data.

A possibility could be to leverage *temporal domain adaptation* to both enable the simultaneous use of reference data from multiple years in each annual land cover mapping task and cope with seasonality shifts in satellite image time series from different years (Capliez et al., 2023). Another relevant possibility worth mentioning would be the incorporation of 2D-CNN models for enhancing the spatial quality of provided maps, including field boundaries delimitation. Yet, such solution would imply a significant increase in the need for reference data, which may render the approach simply unfeasible in areas with limited accessibility like the one studied in this thesis.

7.5 Conclusion

In this thesis we have illustrated different manners to detect *fallow* practice in West African smallholder agriculture. All throughout our exploratory approach, we could provide several methodological clues which may be foundational for the development of an operational strategy for fallow mapping. Furthermore, provided results over a representative study site for the sub-humid Sudanian agricultural systems seem to indicate, with all the due precautions, a non-negligible presence of agricultural land which is not actively cropped during the growing season. Due to the technical limitations in terms of data availability, it is not possible to determine to what extent this land might correspond unquestionably to a regular implementation of some type of *fallow* practice. Yet, these

insights provide evidence of the presence of a highly dynamic agricultural system that cannot be overlooked, and definitely confirm the necessity of pursuing methodological research on the accurate mapping of this practice at broad scale.

Bibliography

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., et al., 2016. {TensorFlow}: a system for {Large-Scale} machine learning, in: 12th USENIX symposium on operating systems design and implementation (OSDI 16), pp. 265–283.

Abdoulaye, D., Bruno, B., Bétéo, Z., Hamma, Y., 2017. Impact of climate change on cotton production in Burkina Faso. African Journal of Agricultural Research 12, 494–501. URL: http://academicjournals.org/journal/AJAR/article-abstract/01E4D7A62764, doi:10.5897/AJAR2015.10763.

Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S., 2012. Slic superpixels compared to state-of-the-art superpixel methods. IEEE Transactions on Pattern Analysis and Machine Intelligence 34, 2274–2282. doi:10.1109/TPAMI.2012.120.

Akpo, L.E., Masse, D., Grouzis, M., 2002. Durée de jachère et valeur pastorale de la végétation herbacée en zone soudanienne au Sénégal. Revue d'élevage et de médecine vétérinaire des pays tropicaux 55, 275. URL: http://revues.cirad.fr/index.php/REMVT/article/view/9815, doi:10.19182/remvt.9815.

Baatz, M., Schäpe, A., Strobl, J., Blaschke, T., Griesebner, G., 2000. Multiresolution segmentation-an optimization approach for high quality multi-scale image segmentation. angewandte geographische informationsverarbeitung, 12, 12–23. retrieved from internal-pdf. xn–baatz_schpe_2000-3891068462-jkc/Baatz_Sch .

Barnes, E.M., Clarke, T.R., Richards, S.E., Colaizzi, P.D., Haberland, J., Kostrzewski, M., Waller, P.M., Choi, C.Y., Riley, E., Thompson, T.L., Lascano, R.J., Li, H., Moran, M.S., Robert, P.C., Rust, R.H., Larson, W.E., 2000. Coincident detection of crop water stress, nitrogen status and canopy density using ground-based multispectral data., in: Proceedings of the 5th International Conference on Precision Agriculture and other resource management July 16-19, 2000, Bloomington, MN USA.

Bekker, J., Davis, J., 2020. Learning from positive and unlabeled data: A survey. Machine Learning 109, 719–760.

Bellman, R., 1966. Dynamic programming. Science 153, 34–37.

Boffa, J.M., 1999. Agroforestry parklands in sub-Saharan Africa. 34, Food & Agriculture Org.

Bontemps, S., Defournya, P., Van Bogaert, E., Weber, J.L., Arino, O., 2009. GlobCorine–A joint EEA-ESA project for operational land dynamics monitoring at pan-European scale, in: The 33rd international symposium on remote sensing of environment. Tex.organization: Tucson/Arizona USA.

Bottou, L., 2010. Large-scale machine learning with stochastic gradient descent, in: Proceedings of COMPSTAT'2010: 19th International Conference on Computational StatisticsParis France, August 22-27, 2010 Keynote, Invited and Contributed Papers, Springer. pp. 177–186.

Breiman, L., 2001. Random forests. Machine learning 45, 5–32.

Brown, C.F., Brumby, S.P., Guzder-Williams, B., Birch, T., Hyde, S.B., Mazzariello, J., Czerwinski, W., Pasquarella, V.J., Haertel, R., Ilyushchenko, S., Schwehr, K., Weisse, M., Stolle, F., Hanson, C., Guinan, O., Moore, R., Tait, A.M., 2022. Dynamic world, near real-time global 10 m land use land cover mapping. Scientific Data 9, 251. URL: https://www.nature.com/articles/s41597-022-01307-4, doi:10.1038/s41597-022-01307-4.

Buchhorn, M., Smets, B., Bertels, L., Roo, B.D., Lesiv, M., Tsendbazar, N.E., Herold, M., Fritz, S., 2020. Copernicus Global Land Service: Land Cover 100m: collection 3: epoch 2019: Globe. URL: https://zenodo.org/record/3939050, doi:10.5281/zenodo.3939050. type: dataset.

Bégué, A., Arvor, D., Bellon, B., Betbeder, J., de Abelleyra, D., P. D. Ferraz, R., Lebourgeois, V., Lelong, C., Simões, M., R. Verón, S., 2018. Remote Sensing and Cropping Practices: A Review. Remote Sensing 10, 99. URL: http://www.mdpi.com/2072-4292/10/1/99, doi:10.3390/rs10010099.

Capliez, E., Ienco, D., Gaetano, R., Baghdadi, N., Salah, A.H., 2023. Temporal-domain adaptation for satellite image time-series land-cover mapping with adversarial learning and spatially aware self-training. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 16, 3645–3675.

Castro Alvarado, E., 2022. Fallow ID: Characterization and mapping of fallow fields in West-Africa study case using Sentinel-2. URL: https://earth.esa.int/living-planet-symposium-2022-presentations/.

Castro Alvarado, E., Bégué, A., Leroux, L., Gaetano, R., 2023. A multi-year land use trajectory strategy for non-active agricultural land mapping in sub-humid west africa. International Journal of Applied Earth Observation and Geoinformation 122, 103398.

Chollet, F., et al., 2015. Keras. URL: https://github.com/fchollet/keras.

Dara, A., Baumann, M., Kuemmerle, T., Pflugmacher, D., Rabe, A., Griffiths, P., Hölzel, N., Kamp, J., Freitag, M., Hostert, P., 2018. Mapping the timing of cropland abandonment and recultivation in northern Kazakhstan using annual Landsat time series. Remote Sensing of Environment 213, 49–60. URL: https://linkinghub.elsevier.com/retrieve/pii/S0034425718302268, doi:10.1016/j.rse.2018.05.005.

Dayamba, S.D., Djoudi, H., Zida, M., Sawadogo, L., Verchot, L., 2016. Biodiversity and carbon stocks in different land use types in the Sudanian Zone of Burkina Faso, West Africa. Agriculture, Ecosystems & Environment 216, 61–72. URL: https://linkinghub.elsevier.com/retrieve/pii/S016788091530089X, doi:10.1016/j.agee.2015.09.023.

Desa, U., others, 2016. Transforming our world: The 2030 agenda for sustainable development. URL: https://sdgs.un.org/2030agenda.

Didan, K., Huete, A., SIPS, M., 2015. MOD13Q1 MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid. http://doi.org/10.5067/MODIS/MOD13Q1.006.

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., Bargellini, P., 2012. Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. Remote Sensing of Environment 120, 25–36. URL: https://www.sciencedirect.com/science/article/pii/S0034425712000636, doi:10.1016/j.rse.2011.11.026.

Ebel, P., Meraner, A., Schmitt, M., Zhu, X.X., 2020. Multisensor data fusion for cloud removal in global and all-season sentinel-2 imagery. IEEE Transactions on Geoscience and Remote Sensing 59, 5866–5878.

Escadafal, R., Girard, M.C., Courault, D., 1989. Munsell soil color and soil reflectance in the visible spectral bands of landsat MSS and TM data. Remote Sensing of Environment 27, 37–46. URL: https://www.sciencedirect.com/science/article/pii/0034425789900357, doi:10.1016/0034-4257(89)90035-7.

Estel, S., Kuemmerle, T., Alcántara, C., Levers, C., Prishchepov, A., Hostert, P., 2015. Mapping farmland abandonment and recultivation across Europe using MODIS NDVI time series. Remote Sensing of Environment 163, 312–325. URL: https://linkinghub.elsevier.com/retrieve/pii/S003442571500125X, doi:10.1016/j.rse.2015.03.028.

Estel, S., Kuemmerle, T., Levers, C., Baumann, M., Hostert, P., 2016. Mapping cropland-use intensity across Europe using MODIS NDVI time series. Environmental Research Letters 11, 024015. URL: https://dx.doi.org/10.1088/1748-9326/11/2/024015, doi:10.1088/1748-9326/11/2/024015. publisher: IOP Publishing.

FAO, 2022a. FAOSTAT statistical database. URL: https://search.library.wisc.edu/catalog/999890171702121.

FAO, 2022b. Term Portal. FAOTERM. URL: https://www.fao.org/faoterm/en/.

FAO, I., 2023. The State of Food Security and Nutrition in the World 2023: Urbanization, agrifood systems transformation and healthy diets across the rural–urban continuum. Number 2023 in The State of Food Security and Nutrition in the World (SOFI), FAO, IFAD, UNICEF, WFP, WHO, Rome, Italy. URL: https://www.fao.org/documents/card/en/c/cc3017en, doi:10.4060/cc3017en.

Faye, J.B., Hopple, A.M., Bridgham, S.D., 2021. Indigenous farming practices increase millet yields in Senegal, West Africa. Agroecology and Sustainable Food Systems 45, 159–174. URL: https://doi.org/10.1080/21683565.2020.1815927, doi:10.1080/21683565.2020.1815927. publisher: Taylor & Francis _eprint: https://doi.org/10.1080/21683565.2020.1815927.

Floret, C., Pontanier, R., 2001. La jachère en Afrique tropicale. volume 2. John Libbey Eurotext, Paris.

Funk, C.C., Peterson, P.J., Landsfeld, M.F., Pedreros, D.H., Verdin, J.P., Rowland, J.D., Romero, B.E., Husak, G.J., Michaelsen, J.C., Verdin, A.P., others, 2014. A quasi-global precipitation time series for drought monitoring. US Geological Survey data series 832, 1–12. Publisher: US Geological Survey Reston, VA, USA.

Gaetano, R., Dupuy, S., Lebourgeois, V., Le Maire, G., Tran, A., Jolivot, A., Bégué, A., 2019. The MORINGA processing chain: automatic object-based land cover classification of tropical agrosystems using multi-sensor satellite imagery, in: Living Planet Symposium, LPS 2019. URL: https://agritrop.cirad.fr/594650/. italian Space Agency.

Gaetano, R., Scarpa, G., Poggi, G., 2007. A hierarchical segmentation algorithm for multiresolution satellite images, in: 2007 IEEE International Geoscience and Remote Sensing Symposium, pp. 1885–1888. doi:10.1109/IGARSS.2007.4423192.

Gao, B.c., 1996. NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. Remote Sensing of Environment 58, 257–266. URL: https://www.sciencedirect.com/science/article/pii/S0034425796000673, doi:10.1016/S0034-4257(96)00067-3.

Gao, F., Anderson, M.C., Zhang, X., Yang, Z., Alfieri, J.G., Kustas, W.P., Mueller, R., Johnson, D.M., Prueger, J.H., 2017. Toward mapping crop progress at field scales through fusion of landsat and modis imagery. Remote Sensing of Environment 188, 9–25.

García, M.L., Caselles, V., 1991. Mapping burns and natural reforestation using thematic Mapper data. Geocarto International 6, 31–37. URL: https://doi.org/10.1080/10106049109354290, doi:10.1080/10106049109354290. publisher: Taylor & Francis _eprint: https://doi.org/10.1080/10106049109354290.

Gbodjo, Y.J.E., Ienco, D., Leroux, L., 2021. Benchmarking statistical modelling approaches with multi-source remote sensing data for millet yield monitoring: a case study of the groundnut basin in central senegal. International Journal of Remote Sensing 42, 9285–9308.

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep learning. MIT press.

Graef, F., Haigis, J., 2001. Spatial and temporal rainfall variability in the sahel and its effects on farmers' management strategies. Journal of Arid Environments 48, 221–231.

Hoover, A., Jean-Baptiste, G., Jiang, X., Flynn, P., Bunke, H., Goldgof, D., Bowyer, K., Eggert, D., Fitzgibbon, A., Fisher, R., 1996. An experimental comparison of range image segmentation algorithms. IEEE Transactions on Pattern Analysis and Machine Intelligence 18, 673–689. doi:10.1109/34.506791. conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence.

Inglada, J., 2017. Phenological analysis for image time series, phenootb. https://gitlab.orfeo-toolbox.org/jinglada/phenotb [Accessed: 2023-10-19].

Inglada, J., Arias, M., Tardy, B., Hagolle, O., Valero, S., Morin, D., Dedieu, G., Sepulcre, G., Bontemps, S., Defourny, P., Koetz, B., 2015. Assessment of an Operational System for Crop Type Map Production Using High Temporal and Spatial Resolution Satellite Optical Imagery. Remote Sensing 7, 12356–12379. URL: https://www.mdpi.com/2072-4292/7/9/12356, doi:10.3390/rs70912356. number: 9 Publisher: Multidisciplinary Digital Publishing Institute.

Inglada, J., Christophe, E., 2009. The orfeo toolbox remote sensing image processing software, in: 2009 IEEE International Geoscience and Remote Sensing Symposium, IEEE. pp. IV–733.

Inglada, J., Vincent, A., Arias, M., Tardy, B., Morin, D., Rodes, I., 2017. Operational High Resolution Land Cover Map Production at the Country Scale Using Satellite Image Time Series. Remote Sensing 9, 95. URL: https://www.mdpi.com/2072-4292/9/1/95, doi:10.3390/rs9010095. number: 1 Publisher: Multidisciplinary Digital Publishing Institute.

Jahel, C., Vall, E., Rodriguez, Z., Bégué, A., Baron, C., Augusseau, X., Lo Seen, D., 2018. Analysing plausible futures from past patterns of land change in West Burkina Faso. Land Use Policy 71, 60–74. URL: https://linkinghub.elsevier.com/retrieve/pii/S026483771631081X, doi:10.1016/j.landusepol.2017.11.025.

Jolivot, A., Lebourgeois, V., Ameline, M., Andriamanga, V., Bellón, B., Castets, M., Crespin-Boucaud, A., Defourny, P., Diaz, S., Dieye, M., Dupuy, S., Ferraz, R., Gaetano, R., Gely, M., Jahel, C., Kabore, B., Lelong, C., Le Maire, G., Leroux, L., Lo Seen, D., Muthoni, M., Ndao, B., Newby, T., De Oliveira Santos, C.L.M., Rasoamalala, E., Simoes, M., Thiaw, I., Timmermans, A., Tran, A., Bégué, A., 2021. Harmonized in situ JECAM datasets for agricultural land use mapping and monitoring in tropical countries. Earth System Science Data Discussions 2021, 1–22. URL: https://essd.copernicus.org/preprints/essd-2021-125/, doi:10.5194/essd-2021-125.

Jouve, P., 1993. Usages et fonctions de la jachère dans les systèmes de production d'afrique tropicale et du maghreb. Cahiers Agricultures 2, 308–317.

Kaya, B., Nair, P.K.R., 2001. Soil fertility and crop yields under improved-fallow systems in southern Mali. Agroforestry Systems 52, 1–11. URL: https://doi.org/10.1023/A:1010717629129, doi:10.1023/A:1010717629129.

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 Sensing of Environment 114, 2897–2910.

Lab, F.C., for International Earth Science Information Network-CIESIN-Columbia University, C., 2016. High resolution settlement layer (hrsl). Source imagery for HRSL© 2016 DigitalGlobe .

Lasko, K., 2022. Gap filling cloudy sentinel-2 ndvi and ndwi pixels with multi-frequency denoised c-band and l-band synthetic aperture radar (sar), texture, and shallow learning techniques. Remote Sensing 14, 4221.

Lebel, T., Le Barbé, L., 1997. Rainfall monitoring during hapex-sahel. J. General rainfall conditions and climatology, l Hydro , 88–189.

Lericollais, A., Milleville, P., 1993. La jachère dans les systèmes agro-pastoraux sereer au sénégal. La jachere en Afrique de l'Ouest , 133–145.

Leroux, L., Jolivot, A., Bégué, A., Seen, D.L., Zoungrana, B., 2014. How Reliable is the MODIS Land Cover Product for Crop Mapping Sub-Saharan Agricultural Landscapes? Remote Sensing 6, 8541–8564. URL: https://www.mdpi.com/2072-4292/6/9/8541, doi:10.3390/rs6098541. number: 9 Publisher: Multidisciplinary Digital Publishing Institute.

Lowder, S.K., Skoet, J., Raney, T., 2016. The number, size, and distribution of farms, smallholder farms, and family farms worldwide. World development 87, 16–29.

Manlay, R.J., Cadet, P., Thioulouse, J., Chotte, J.L., 2000a. Relationships between abiotic and biotic soil properties during fallow periods in the sudanian zone of Senegal. Applied Soil Ecology 14, 89–101. URL: https://linkinghub.elsevier.com/retrieve/pii/S0929139300000524, doi:10.1016/S0929-1393(00)00052-4.

Manlay, R.J., Cadet, P., Thioulouse, J., Chotte, J.L., 2000b. Relationships between abiotic and biotic soil properties during fallow periods in the sudanian zone of Senegal. Applied Soil Ecology 14, 89–101. URL: https://linkinghub.elsevier.com/retrieve/pii/S0929139300000524, doi:10.1016/S0929-1393(00)00052-4.

Mathieu, R., Pouget, M., Cervelle, B., Escadafal, R., 1998. Relationships between Satellite-Based Radiometric Indices Simulated Using Laboratory Reflectance Data and Typic Soil Color of an Arid Environment. Remote Sensing of Environment 66, 17–28. URL: https://www.sciencedirect.com/science/article/pii/S0034425798000303, doi:10.1016/S0034-4257(98)00030-3.

McFeeters, S.K., 1996. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. International Journal of Remote Sensing 17, 1425–1432. URL: https://doi.org/10.1080/01431169608948714, doi:10.1080/01431169608948714. publisher: Taylor & Francis _eprint: https://doi.org/10.1080/01431169608948714.

Meraner, A., Ebel, P., Zhu, X.X., Schmitt, M., 2020. Cloud removal in sentinel-2 imagery using a deep residual neural network and sar-optical data fusion. ISPRS Journal of Photogrammetry and Remote Sensing 166, 333–346.

Miller, M.A.E., Shepherd, K.D., Kisitu, B., Collinson, J., 2021. iSDAsoil: The first continent-scale soil property map at 30 m resolution provides a soil information revolution for Africa. PLOS Biology 19, e3001441. URL: https://dx.plos.org/10.1371/journal.pbio.3001441, doi:10.1371/journal.pbio.3001441.

Müller, A.C., Guido, S., 2016. Introduction to machine learning with Python: a guide for data scientists. " O'Reilly Media, Inc.".

Ouedraogo, I., Tigabu, M., Savadogo, P., Compaoré, H., Odén, P.C., Ouadba, J.M., 2010. Land cover change and its relation with population dynamics in Burkina Faso, West Africa: LAND COVER CHANGE IN BURKINA FASO. Land Degradation & Development 21, 453–462. URL: https://onlinelibrary.wiley.com/doi/10.1002/ldr.981, doi:10.1002/ldr.981.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research 12, 2825–2830.

Pelletier, C., Webb, G.I., Petitjean, F., 2019. Temporal Convolutional Neural Network for the Classification of Satellite Image Time Series. Remote Sensing 11, 523. URL: https://www.mdpi.com/2072-4292/11/5/523, doi:10.3390/rs11050523. number: 5 Publisher: Multidisciplinary Digital Publishing Institute.

Ploton, P., Mortier, F., Réjou-Méchain, M., Barbier, N., Picard, N., Rossi, V., Dormann, C., Cornu, G., Viennois, G., Bayol, N., Lyapustin, A., Gourlet-Fleury, S., Pélissier,

R., 2020. Spatial validation reveals poor predictive performance of large-scale ecological mapping models. Nature Communications 11. URL: https://doi.org/10.1038/s41467-020-18321-y, doi:10.1038/s41467-020-18321-y.

Qiu, B., Lin, D., Chen, C., Yang, P., Tang, Z., Jin, Z., Ye, Z., Zhu, X., Duan, M., Huang, H., Zhao, Z., Xu, W., Chen, Z., 2022. From cropland to cropped field: A robust algorithm for national-scale mapping by fusing time series of Sentinel-1 and Sentinel-2. International Journal of Applied Earth Observation and Geoinformation 113, 103006. URL: https://www.sciencedirect.com/science/article/pii/S1569843222001947, doi:10.1016/j.jag.2022.103006.

Quillemin, R., 1956. Evolution de l'agriculture autochtone dans les savanes de l'oubangui. L'Agronomie tropicale 11, 39–61.

de Ridder, N., Breman, H., van Keulen, H., Stomph, T.J., 2004. Revisiting a 'Cure against land hunger': Soil fertility management and farming systems dynamics in the West African Sahel. Agricultural Systems 80, 109–131. URL: https://linkinghub.elsevier.com/retrieve/pii/S0308521X0300101X, doi:10.1016/j.agsy.2003.06.004. tex.options: useprefix=true.

Ringius, L., 2002. Soil Carbon Sequestration and the CDM: Opportunities and Challenges for Africa , 25.

Rouse, J.W., Haas, R.H., Deering, D.W., Schell, J.A., Harlan, J.C., 1974. Monitoring the Vernal Advancement and Retrogradation (Green Wave Effect) of Natural Vegetation. Technical Report E75-10354. Texas A&M University. URL: https://ntrs.nasa.gov/citations/19750020419. nTRS Author Affiliations: Texas A&M Univ.

Rufin, P., Bey, A., Picoli, M., Meyfroidt, P., 2022. Large-area mapping of active cropland and short-term fallows in smallholder landscapes using PlanetScope data. International Journal of Applied Earth Observation and Geoinformation 112, 102937. URL: https://linkinghub.elsevier.com/retrieve/pii/S1569843222001340, doi:10.1016/j.jag.2022.102937.

Ruthenberg, H., 1971. Farming systems in the tropics. Clarendon Press - Oxford.

Samaké, O., Smaling, E., Kropff, M., Stomph, T., Kodio, A., 2005. Effects of cultivation practices on spatial variation of soil fertility and millet yields in the Sahel of Mali. Agriculture, Ecosystems & Environment 109, 335–345. URL: https://linkinghub.elsevier.com/retrieve/pii/S0167880905001246, doi:10.1016/j.agee.2005.02.024.

Samasse, K., Hanan, N., Tappan, G., Diallo, Y., 2018. Assessing Cropland Area in West Africa for Agricultural Yield Analysis. Remote Sensing 10, 1785. URL: http://www.mdpi.com/2072-4292/10/11/1785, doi:10.3390/rs10111785.

dos Santos, E.P., da Silva, D.D., do Amaral, C.H., Fernandes-Filho, E.I., Dias, R.L.S., 2022. A machine learning approach to reconstruct cloudy affected vegetation indices imagery via data fusion from sentinel-1 and landsat 8. Computers and Electronics in Agriculture 194, 106753.

Scarpa, G., Gargiulo, M., Mazza, A., Gaetano, R., 2018. A cnn-based fusion method for feature extraction from sentinel data. Remote Sensing 10, 236.

Snapp, S., Rahmanian, M., Batello, C., Calles, T., 2018. Pulse crops for sustainable farms in sub-saharan africa. FAO. doi:10.18356/6795bfaf-en.

Tittonell, P., Giller, K.E., 2013. When yield gaps are poverty traps: The paradigm of ecological intensification in African smallholder agriculture. Field Crops Research 143, 76–90. URL: https://www.sciencedirect.com/science/article/pii/S0378429012003346, doi:10.1016/j.fcr.2012.10.007.

Tong, X., Brandt, M., Hiernaux, P., Herrmann, S., Rasmussen, L.V., Rasmussen, K., Tian, F., Tagesson, T., Zhang, W., Fensholt, R., 2020. The forgotten land use class: Mapping of fallow fields across the Sahel using Sentinel-2. Remote Sensing of Environment 239, 111598. URL: https://linkinghub.elsevier.com/retrieve/pii/S0034425719306182, doi:10.1016/j.rse.2019.111598.

Tong, X., Brandt, M., Vang Rasmussen, L., Hiernaux, P., Bech Bruun, T., Reiner, F., M. Abdi, A., M. Herrmann, S., Li, S., Fensholt, R., 2022. Nano-satellites uphold Boserup's theory of smallholder agricultural intensification. URL: https://www.researchsquare.com, doi:10.21203/rs.3.rs-2041995/v1.

Verbesselt, J., Hyndman, R., Newnham, G., Culvenor, D., 2010. Detecting trend and seasonal changes in satellite image time series. Remote sensing of Environment 114, 106–115.

Vollset, S.E., Goren, E., Yuan, C.W., Cao, J., Smith, A.E., Hsiao, T., Bisignano, C., Azhar, G.S., Castro, E., Chalek, J., Dolgert, A.J., Frank, T., Fukutaki, K., Hay, S.I., Lozano, R., Mokdad, A.H., Nandakumar, V., Pierce, M., Pletcher, M., Robalik, T., Steuben, K.M., Wunrow, H.Y., Zlavog, B.S., Murray, C.J.L., 2020. Fertility, mortality, migration, and population scenarios for 195 countries and territories from 2017 to 2100: a forecasting analysis for the Global Burden of Disease Study. The Lancet 396, 1285–1306. URL: https://www.sciencedirect.com/science/article/pii/S0140673620306772, doi:10.1016/S0140-6736(20)30677-2.

Wallace, C.S., Thenkabail, P., Rodriguez, J.R., Brown, M.K., 2017. Fallow-land Algorithm based on Neighborhood and Temporal Anomalies (FANTA) to map planted *versus* fallowed croplands using MODIS data to assist in drought studies leading to water and food security assessments. GIScience & Remote Sensing 54, 258–282. URL: https://www.tandfonline.com/doi/full/10.1080/15481603.2017.1290913, doi:10.1080/15481603.2017.1290913.

Wu, J., Lin, L., Li, T., Cheng, Q., Zhang, C., Shen, H., 2022. Fusing landsat 8 and sentinel-2 data for 10-m dense time-series imagery using a degradation-term constrained deep network. International Journal of Applied Earth Observation and Geoinformation 108, 102738.

Wu, Z., Thenkabail, P.S., Mueller, R., Zakzeski, A., Melton, F., Johnson, L., Rosevelt, C., Dwyer, J., Jones, J., Verdin, J.P., 2014. Seasonal cultivated and fallow cropland mapping using MODIS-based automated cropland classification algorithm. Journal of Applied Remote Sensing 8, 083685. URL: https://www.spiedigitallibrary.org/journals/journal-of-applied-remote-sensing/volume-8/issue-1/083685/Seasonal-cultivated-and-fallow-cropland-mapping-using-MODIS-based-automated/10.1117/1.JRS.8.083685.full, doi:10.1117/1.JRS.8.083685. publisher: SPIE.

Xu, H., 2006. Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. International Journal of Remote Sensing 27, 3025–3033. URL: https://doi.org/10.1080/01431160600589179, doi:10.1080/01431160600589179. publisher: Taylor & Francis _eprint: https://doi.org/10.1080/01431160600589179.

Xu, Y., Yu, L., Feng, D., Peng, D., Li, C., Huang, X., Lu, H., Gong, P., 2019. Comparisons of three recent moderate resolution African land cover datasets: CGLS-LC100, ESA-S2-LC20, and FROM-GLC-Africa30. International Journal of Remote Sensing 40, 6185–6202. URL: https://www.tandfonline.com/doi/full/10.1080/01431161.2019.1587207, doi:10.1080/01431161.2019.1587207.

Yin, H., Prishchepov, A.V., Kuemmerle, T., Bleyhl, B., Buchner, J., Radeloff, V.C., 2018. Mapping agricultural land abandonment from spatial and temporal segmentation of Landsat time series. Remote Sensing of Environment 210, 12–24. URL: https://linkinghub.elsevier.com/retrieve/pii/S0034425718300622, doi:10.1016/j.rse.2018.02.050.

Zhang, L., Zhang, L., Du, B., 2016. Deep learning for remote sensing data: A technical tutorial on the state of the art. IEEE Geoscience and remote sensing magazine 4, 22–40.

Zhang, M., Wu, B., Meng, J., Dong, T., You, X., 2014. Fallow land mapping for better crop monitoring in Huang-Huai-Hai Plain using HJ-1 CCD data. IOP Conference Series: Earth and Environmental Science 17, 012048. URL: https://iopscience.iop.org/article/10.1088/1755-1315/17/1/012048, doi:10.1088/1755-1315/17/1/012048.

Zhang, Q., Yuan, Q., Li, J., Li, Z., Shen, H., Zhang, L., 2020. Thick cloud and cloud shadow removal in multitemporal imagery using progressively spatio-temporal patch group deep learning. ISPRS Journal of Photogrammetry and Remote Sensing 162, 148–160.

Zhao, Z., Wang, J., Wang, L., Rao, X., Ran, W., Xu, C., 2023. Monitoring and analysis of abandoned cropland in the Karst Plateau of eastern Yunnan, China based on Landsat time series images. Ecological Indicators 146, 109828. URL: https://www.sciencedirect.com/science/article/pii/S1470160X22013012, doi:10.1016/j.ecolind.2022.109828.

Zoungrana, I., 1993. Les jachères nord-soudaniennes du Burkina Faso: 1. Analyse de la reconstitution de la végétation herbacée, in: La Jachère en Afrique de l'Ouest, Montpellier (France), 1991.

Annexes

Table A.1: Supervised annual fallow mapping strategies per-class accuracy metric. *RF* and *TCNN* stand for supervised Random Forest classifier and TempCNN classifier trials. *OA*, *UA* and *PA* stand for overall accuracy, user accuracy and producer accuracy respectively. For more detail of approaches see **Chapter 4**.

Trial	Year	OA	Class	UA	PA
RF	2016	0.87	non-vegetated	0.93	0.47
RF	2016		evergreen	0.85	0.68
RF	2016		cropped	0.85	0.98
RF	2016		water	1.00	0.99
RF	2016		fallow	0.29	0.01
RF	2016		nat.herbaceus	0.88	0.88
RF	2017	0.89	non-vegetated	0.96	0.72
RF	2017		evergreen	0.85	0.71
RF	2017		cropped	0.91	0.98
RF	2017		water	1.00	0.99
RF	2017		fallow	0.03	0.00
RF	2017		nat.herbaceus	0.87	0.88
RF	2018	0.90	non-vegetated	0.89	0.62
RF	2018		evergreen	0.88	0.80
RF	2018		cropped	0.90	0.97
RF	2018		water	1.00	1.00
RF	2018		fallow	0.20	0.00
RF	2018		nat.herbaceus	0.89	0.91
RF	2020	0.88	non-vegetated	0.85	0.54
RF	2020		evergreen	0.84	0.74
RF	2020		cropped	0.92	0.98
RF	2020		water	1.00	1.00
RF	2020		fallow	0.11	0.01
RF	2020		nat.herbaceus	0.87	0.89
RF	2021	0.91	non-vegetated	0.91	0.50
RF	2021		evergreen	0.90	0.83
RF	2021		cropped	0.92	0.98
RF	2021		water	1.00	1.00
RF	2021		fallow	0.92	0.10
RF	2021		nat.herbaceus	0.90	0.92
TCNN	2016	0.75	non-vegetated	0.32	0.65
TCNN	2016		evergreen	0.77	0.70
TCNN	2016		cropped	0.94	0.66

Continuation Tab. A.1

Trial	Year	OA	Class	UA	PA
TCNN	2016		water	0.99	1.00
TCNN	2016		fallow	0.07	0.40
TCNN	2016		nat.herbaceus	0.87	0.84
TCNN	2017	0.78	non-vegetated	0.46	0.86
TCNN	2017		evergreen	0.78	0.75
TCNN	2017		cropped	0.97	0.75
TCNN	2017		water	1.00	1.00
TCNN	2017		fallow	0.11	0.71
TCNN	2017		nat.herbaceus	0.88	0.82
TCNN	2018	0.83	non-vegetated	0.51	0.78
TCNN	2018		evergreen	0.79	0.81
TCNN	2018		cropped	0.94	0.83
TCNN	2018		water	0.98	1.00
TCNN	2018		fallow	0.06	0.36
TCNN	2018		nat.herbaceus	0.89	0.87
TCNN	2020	0.85	non-vegetated	0.59	0.75
TCNN	2020		evergreen	0.78	0.79
TCNN	2020		cropped	0.96	0.92
TCNN	2020		water	1.00	1.00
TCNN	2020		fallow	0.19	0.70
TCNN	2020		nat.herbaceus	0.88	0.84
TCNN	2021	0.89	non-vegetated	0.40	0.69
TCNN	2021		evergreen	0.82	0.88
TCNN	2021		cropped	0.96	0.91
TCNN	2021		water	1.00	1.00
TCNN	2021		fallow	0.18	0.63
TCNN	2021		nat.herbaceus	0.91	0.88

Trajectory based NAAL mapping

Table A.2: Annual 4-class management accuracy metrics. *RF* and *TCNN* stand for supervised Random Forest classifier and TempCNN classifier trials. *OA*, *UA* and *PA* stand for overall accuracy, user accuracy and producer accuracy respectively. For more detail of approaches see **Chapter 6**.

Classifier	Year	Class	UA	PA	F1-Score	OA
RF	2016	unmanaged	0.89	0.86	0.86	0.88
RF		non-vegetated	0.96	0.70	0.79	
RF		managed	0.86	0.98	0.91	
RF		evergreen	0.94	0.73	0.82	
RF	2017	unmanaged	0.87	0.86	0.86	0.9
RF		non-vegetated	0.74	0.81	0.75	
RF		managed	0.92	0.98	0.95	
RF		evergreen	0.90	0.75	0.80	
RF	2018	unmanaged	0.93	0.90	0.91	0.92
RF		non-vegetated	0.93	0.83	0.83	
RF		managed	0.91	0.97	0.94	
RF		evergreen	0.92	0.87	0.88	
RF	2019	unmanaged	0.95	0.95	0.95	0.96
RF		non-vegetated	0.96	0.94	0.94	
RF		managed	0.97	0.99	0.98	
RF		evergreen	0.95	0.89	0.91	
RF	2020	unmanaged	0.89	0.93	0.90	0.91
RF		non-vegetated	0.94	0.77	0.82	
RF		managed	0.92	0.99	0.95	
RF		evergreen	0.96	0.76	0.84	
RF	2021	unmanaged	0.91	0.92	0.89	0.92
RF		non-vegetated	0.96	0.78	0.88	
RF		managed	0.93	0.98	0.95	
RF		evergreen	0.93	0.83	0.86	
TCNN	2016	unmanaged	0.87	0.85	0.86	0.88
TCNN		managed	0.90	0.97	0.93	
TCNN		evergreen	0.82	0.71	0.76	
TCNN		non-vegetated	0.91	0.74	0.81	
TCNN	2017	unmanaged	0.90	0.86	0.87	0.9
TCNN		managed	0.94	0.97	0.96	
TCNN		evergreen	0.85	0.82	0.83	
TCNN		non-vegetated	0.75	0.83	0.75	
TCNN	2018	unmanaged	0.92	0.88	0.90	0.91
TCNN		managed	0.91	0.97	0.94	
TCNN		evergreen	0.86	0.87	0.86	
TCNN		non-vegetated	0.92	0.82	0.86	
TCNN	2019	unmanaged	0.94	0.95	0.94	0.95
TCNN		managed	0.97	1.00	0.98	
TCNN		evergreen	0.96	0.81	0.84	
TCNN		non-vegetated	0.95	0.96	0.96	

TCNN	2020	unmanaged	0.91	0.89	0.90	0.91
TCNN		managed	0.94	0.98	0.96	
TCNN		evergreen	0.84	0.78	0.80	
TCNN		non-vegetated	0.89	0.76	0.78	
TCNN	2021	unmanaged	0.92	0.89	0.91	0.92
TCNN		managed	0.94	0.98	0.96	
TCNN		evergreen	0.85	0.86	0.85	
TCNN		non-vegetated	0.90	0.71	0.77	

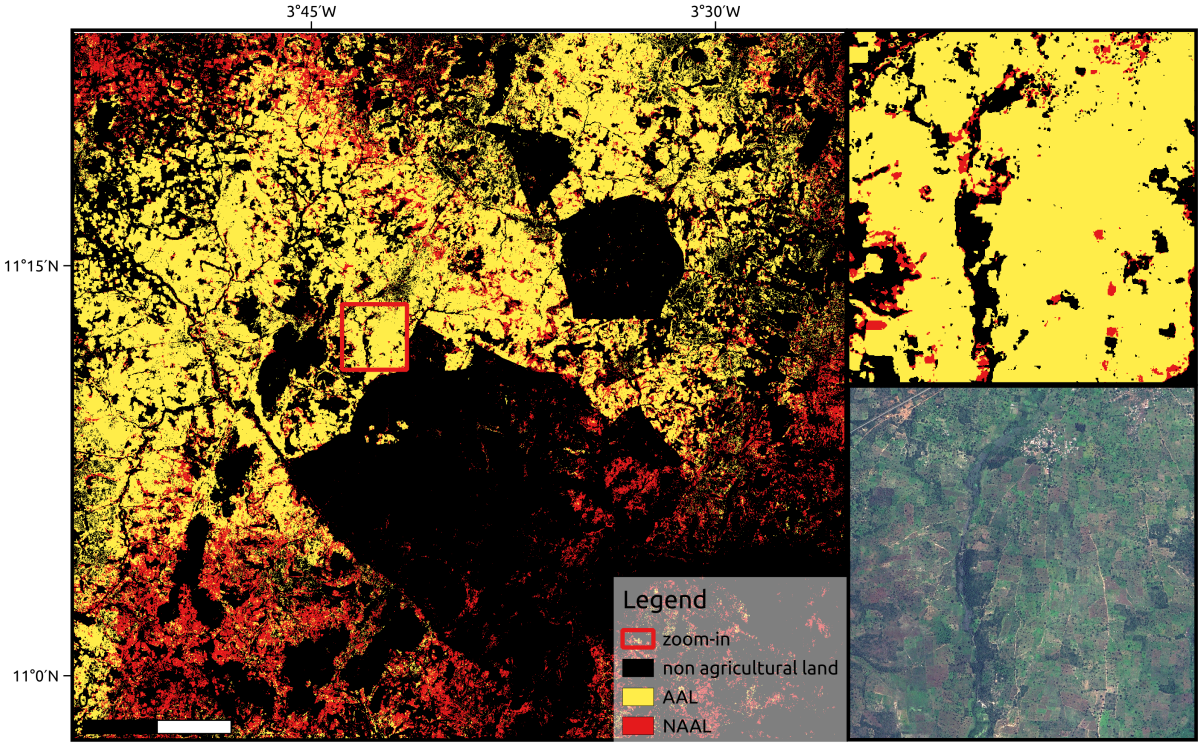


Figure A.1: Annual non-active land (NAAL) map for Koumbia site in 2018, based on Random Forest generated annual management maps. On the right, zoom-in and corresponding SPOT 7 image ©Airbus DS 2018.

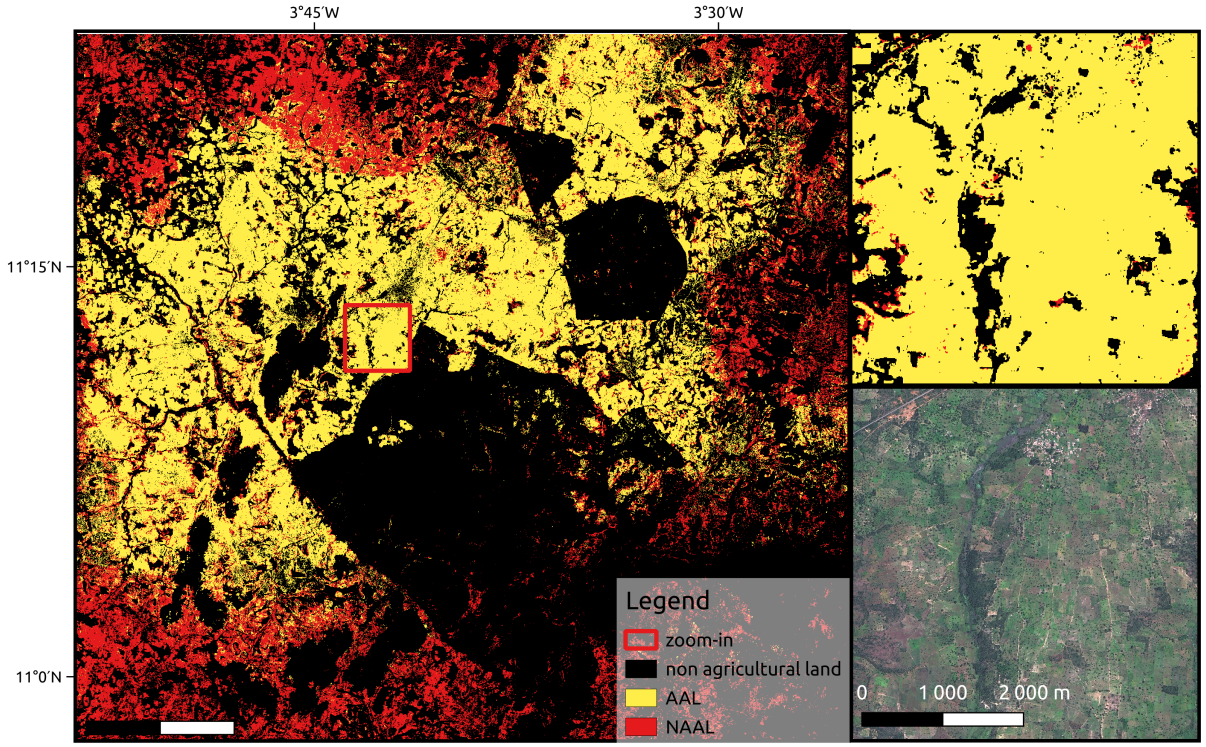


Figure A.2: Annual non-active land (NAAL) map for Koumbia site in 2019, based on Random Forest generated annual management maps. On the right, zoom-in and corresponding SPOT 7 image ©Airbus DS 2019.

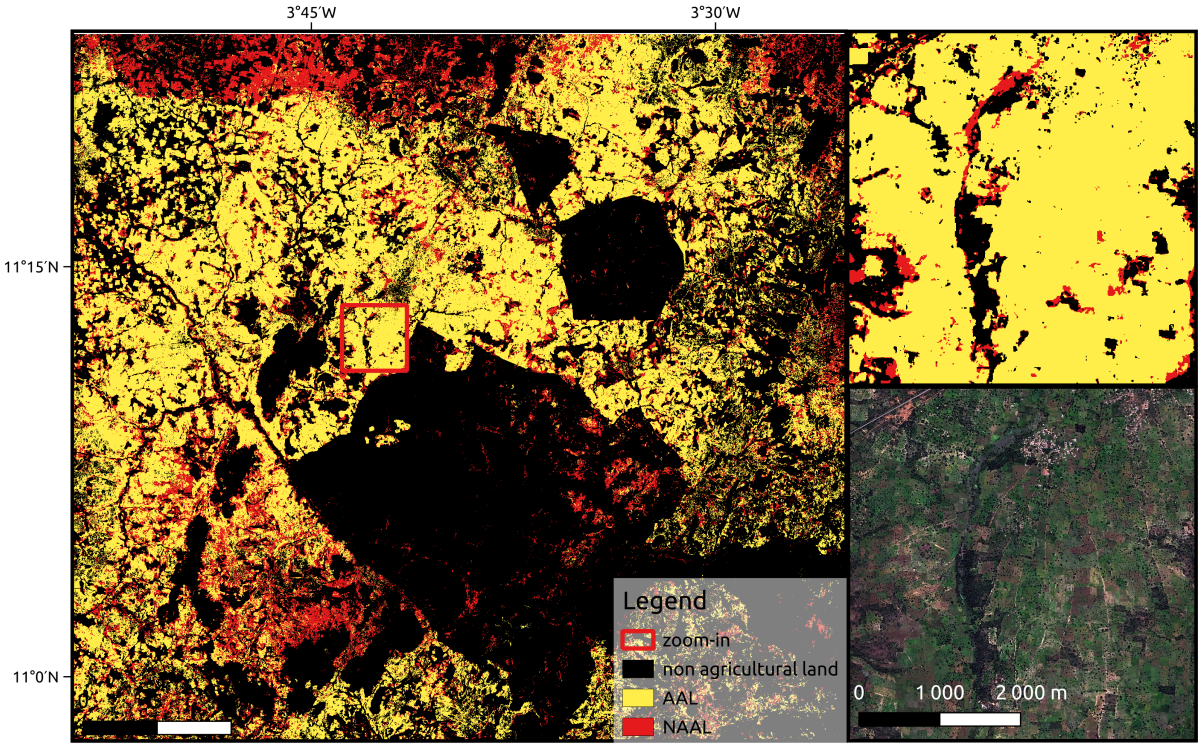


Figure A.3: Annual non-active land (NAAL) map for Koumbia site in 2020, based on Random Forest generated annual management maps. On the right, zoom-in and corresponding SPOT 7 image ©Airbus DS 2020.

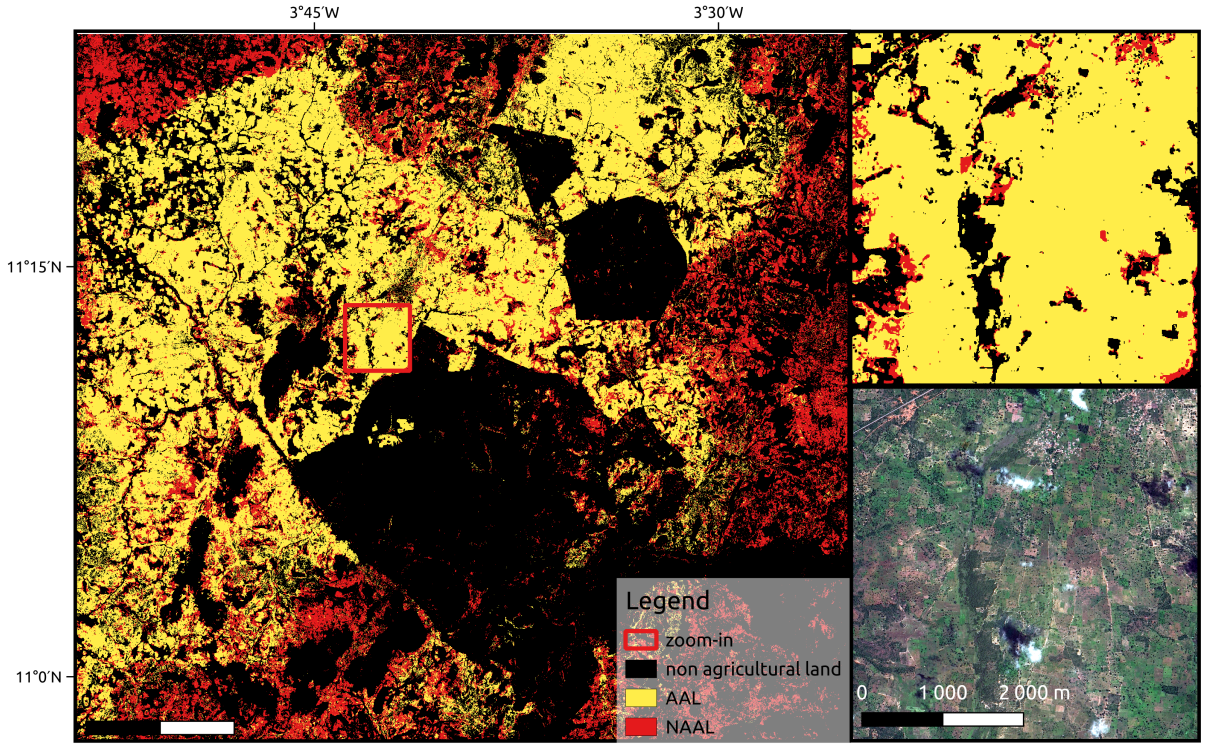


Figure A.4: Annual non-active land (NAAL) map for Koumbia site in 2021, based on Random Forest generated annual management maps. On the right, zoom-in and corresponding SPOT 7 image ©Airbus DS 2021.

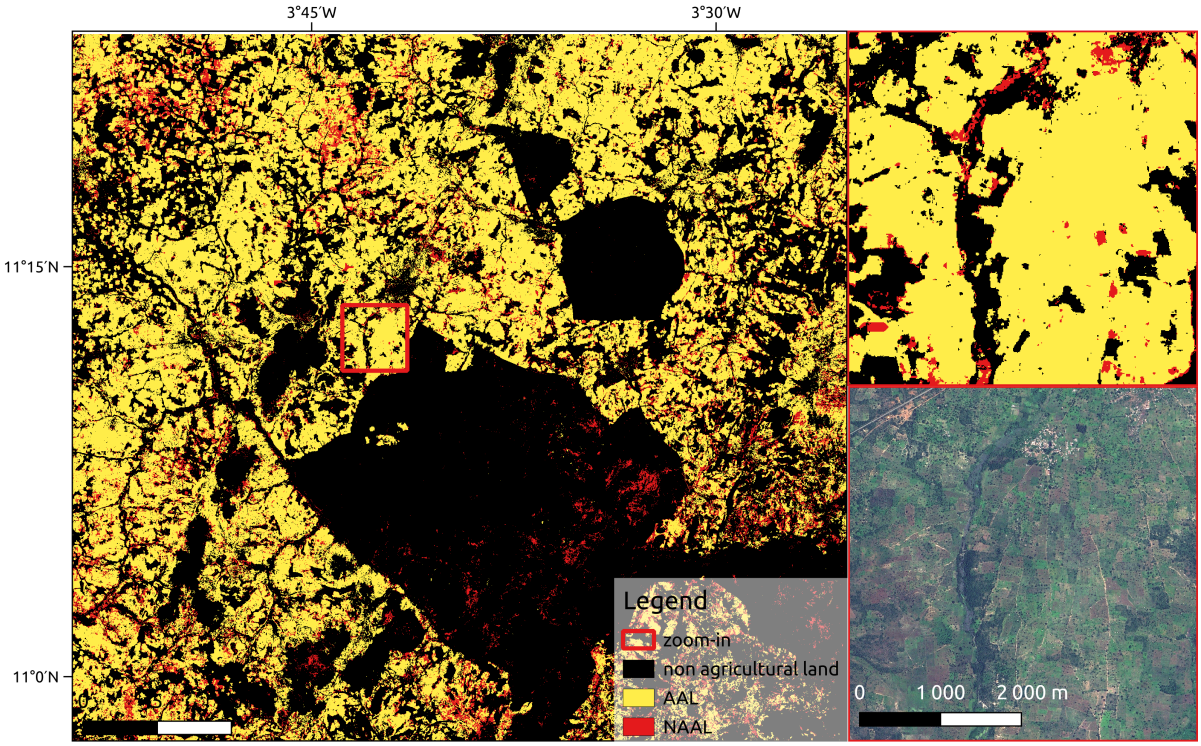


Figure A.5: Annual non-active land (NAAL) map for Koumbia site in 2018, based on TempCNN generated annual management maps. On the right, zoom-in and corresponding SPOT 7 image ©Airbus DS 2018.

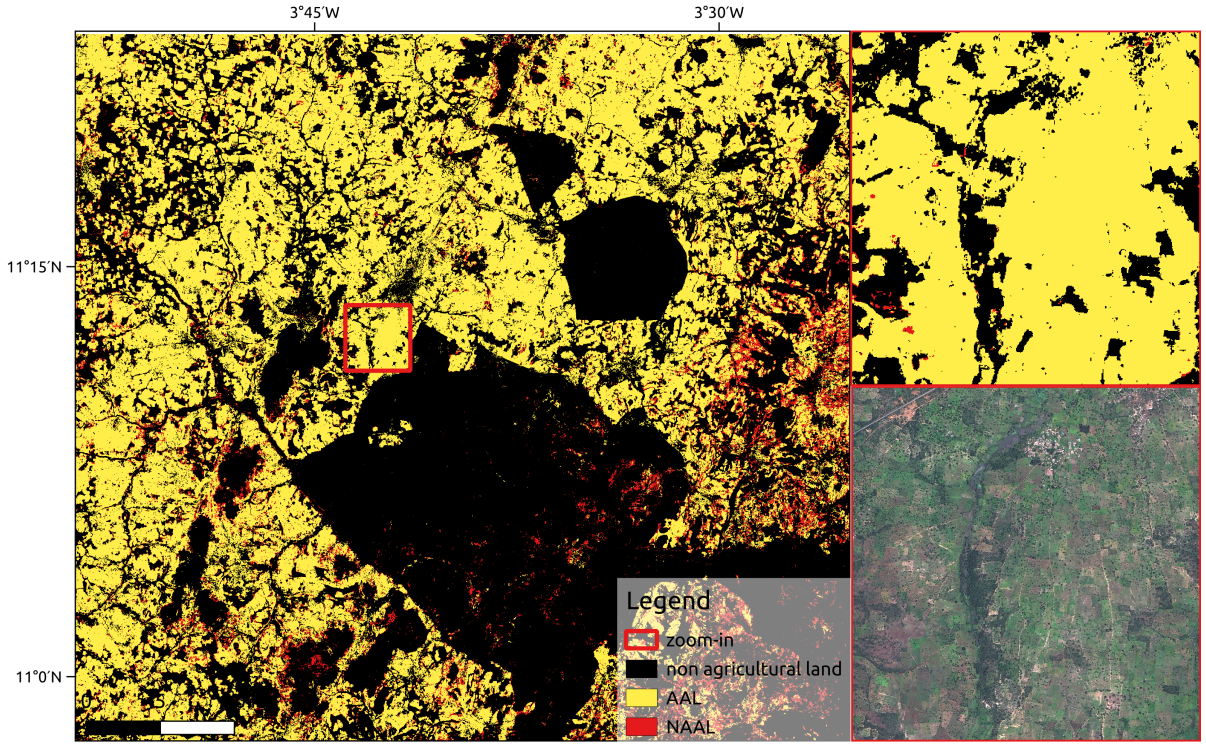


Figure A.6: Annual non-active land (NAAL) map for Koumbia site in 2019, based on TempCNN generated annual management maps. On the right, zoom-in and corresponding SPOT 7 image ©Airbus DS 2019.

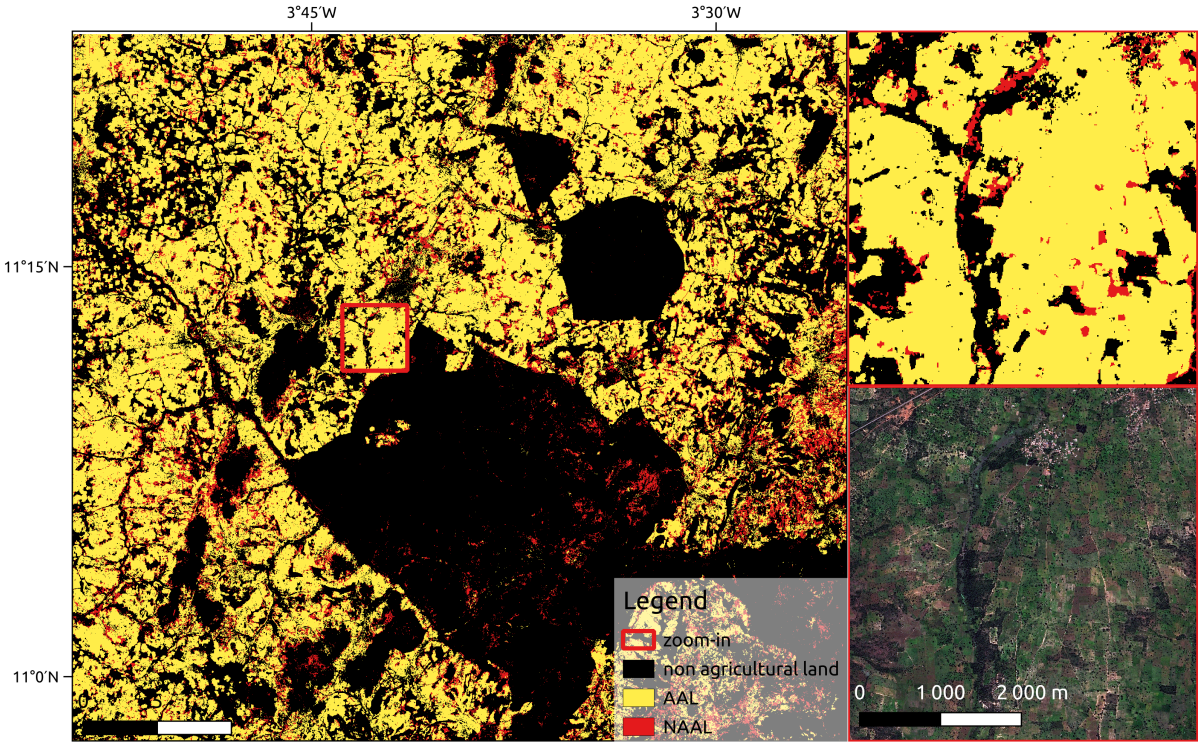


Figure A.7: Annual non-active land (NAAL) map for Koumbia site in 2020, based on TempCNN generated annual management maps. On the right, zoom-in and corresponding SPOT 7 image ©Airbus DS 2020.

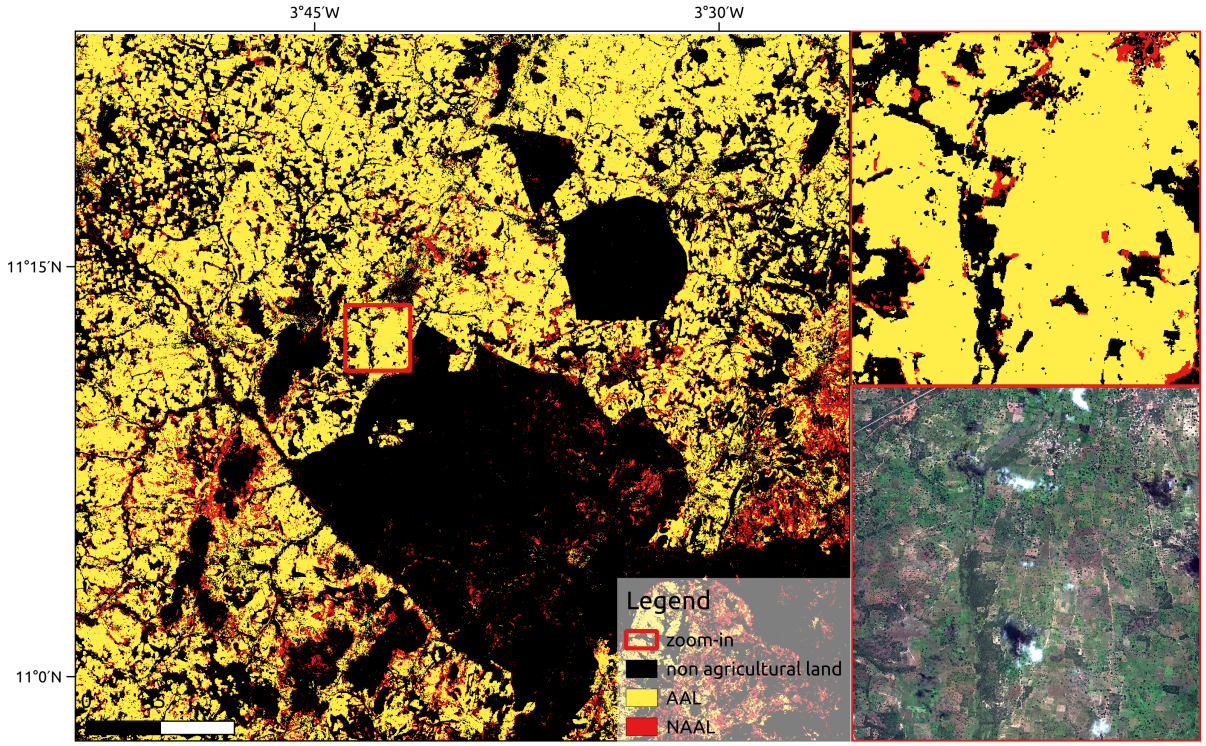


Figure A.8: Annual non-active land (NAAL) map for Koumbia site in 2021, based on TempCNN generated annual management maps. On the right, zoom-in and corresponding SPOT 7 image ©Airbus DS 2021.

Dry season soil map generation procedure

In order to capture the spatial variation linked to soil properties in our study site, we generated a custom soil map classification employing Sentinel-2 imagery. We conducted a clustering process using Orfeo Toolbox's (OTB v7.2.0) Kmeans classification of median values of Sentinel-2 median soil related indices and the Normalized Difference Vegetation Index (NDVI). More specifically, we included the Redness Index (RI), Color Index (CI), Brilliance Index (BI), Brilliance Index II (BI2)(see **Tab. 2.2**) for year 2017 of Sentinel-2 data between the period January-June. During the period chosen (Jan-Jun), there is a reduced amount of cloud coverage and low vegetation density, guarantying better visibility over soil surface. We choose employing a single year of data for avoiding unnecessary data complexity and considered that a single year soil classification is sufficiently representative as it is unlikely to vary across time. Moreover, with the purpose of further enhance soil surface discriminability, a high NDVI mask was employed for reducing Tree influence over the classification process. The specific NDVI threshold for generating the high NDVI mask was set by trial an error to 0.72.

The classification job was conducted setting the number of classes to 5 and generated a first version of dry season soil map (see **Fig. A.9**)

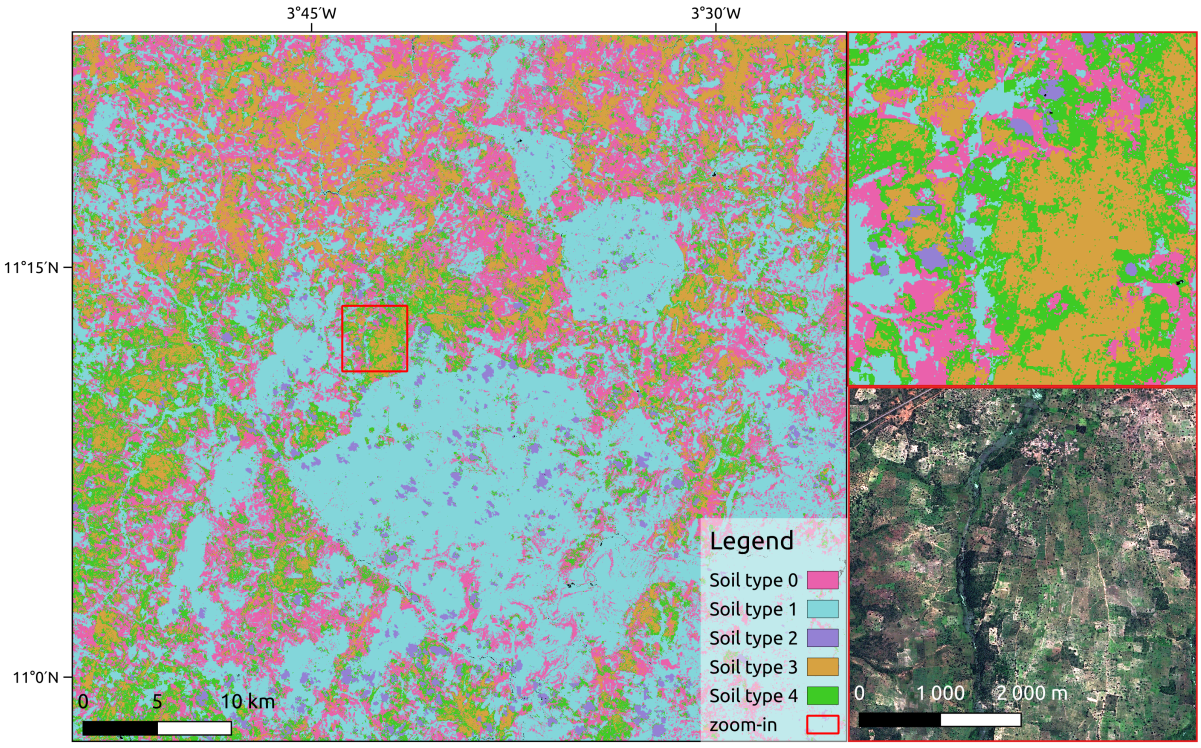


Figure A.9: Dry season 5-K means soil map calculated with Sentinel-2 imagery for period Jan-Jun of year 2017. On the right, zoom-in corresponding to highlighted red area (top) and corresponding SPOT 6 image (bottom) ©Airbus DS 2017.

Resulting map shows thematic sense, based on the fact that large forested areas seem to be associated to a single soil type class (*Soil type 1*). Moreover, soil types spatial distribution shows a slight prevalence of *Soil type 4* on third quadrant, presence of *soil type 3* concentrated in north area and a more sparse *Soil type 0* present in all quadrants. *Soil type 2* is present in small clustered patches accross the entire study site. There is a generalized presence of *salt-and-pepper* noise for the entire study site.

In order to reduce the noise effect, we conducted an object relabelling by soil type majority of a segmentation procedure. The image segmentation was conducted using the OTB's *Generic Region Merging* using *Baatz and Schape* criterion and setting the rest of parameters by test and error for generating compact and visually coherent object, using the median dry season image built in previous step (10 Sentinel-2 bands, NDVI, RI, CI ,BI and BI2) as input. After reclassifying object according to the most prevalent soil type class, we obtained the map presented in **Fig. A.10**. The resulting map was then employed for hypothesis testing in **Chapter 3** (see **Sec. 3.3**).

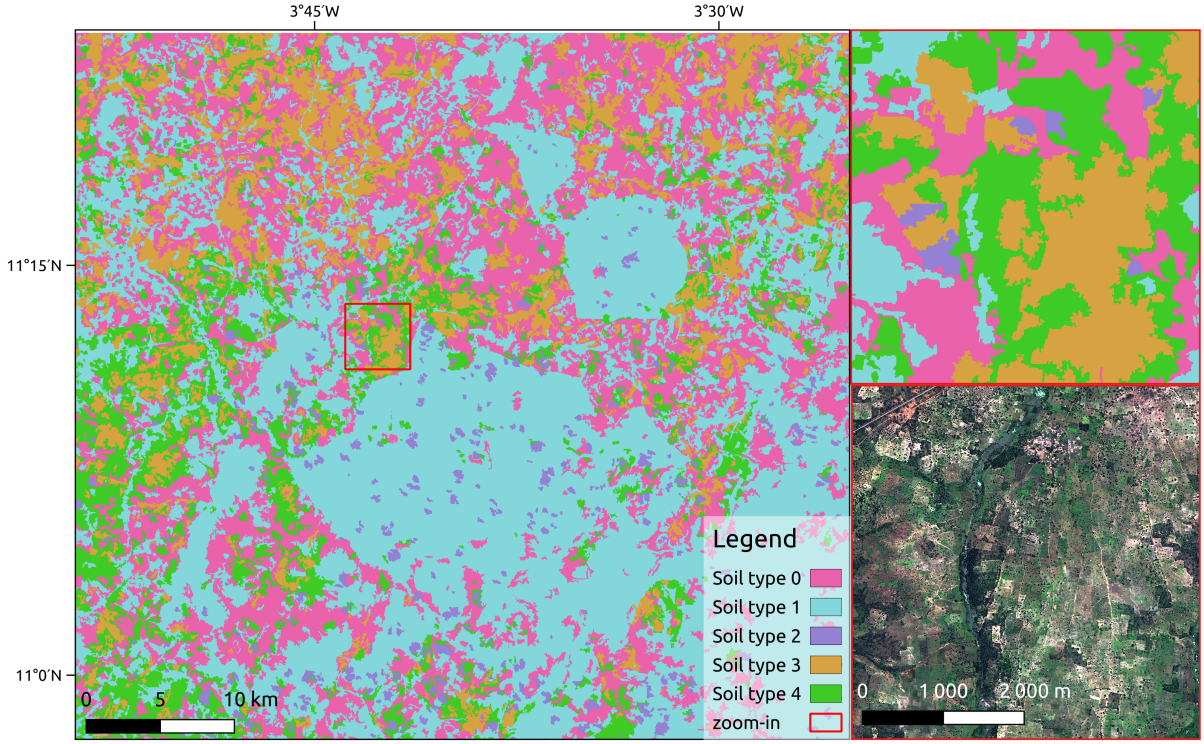


Figure A.10: Processed dry season 5-K means soil map calculated with Sentinel-2 imagery for period Jan-Jun of year 2017. On the right, zoom-in corresponding to highlighted red area (top) and corresponding SPOT 6 image (bottom) ©Airbus DS 2017.

Résumé complet en français

Chapitre 1

L'Afrique subsaharienne (ASS) est depuis des décennies au centre de multiples initiatives, projets et études de recherche visant à éradiquer la faim et la pauvreté et à promouvoir le développement durable dans le sous-continent. La région connaît l'un des taux de croissance démographique les plus élevés au monde. La production agricole est principalement assurée par de petits exploitants, qui opèrent dans des contraintes agro-climatiques et socio-économiques difficiles. Plus précisément, la faiblesse ou l'absence d'intrants minéraux, de mécanisation et d'une agriculture pluviale exercent une pression croissante sur le secteur primaire qui a connu une intensification agricole continue. Cette pression se traduit par l'augmentation de la superficie totale des terres agricoles, ainsi que par le raccourcissement, voire la suppression, des périodes de jachère.

En Afrique subsaharienne, les petits exploitants pratiquent traditionnellement une agriculture extensive où de longues périodes de jachère sont utilisées pour restaurer la fertilité du sol. La littérature relative à la pratique de la jachère montre une grande diversité de mise en oeuvre (voir **Fig. 1.4**). La période d'inactivité (c'est-à-dire la durée de la période non cultivée) est considérée comme l'un des principaux facteurs déterminant les différents états ou typologies des terres en jachère. Sous la pression démographique croissante en Afrique subsaharienne depuis le début des années 60, les auteurs constatent une intensification des pratiques culturales qui se traduit principalement par un raccourcissement des cycles de culture, et donc par une réduction de la durée et de la superficie en jachère.

L'évolution des surfaces en *jachère* est très mal connue en Afrique subsaharienne et dans d'autres régions du Sud global. Récemment, des études ont montré l'intérêt de l'imagerie satellitaire pour cartographier ces surfaces à grande échelle en utilisant au préalable un masque de terres cultivées. Malheureusement, dans les régions dominée par une petite agriculture, les masques de culture sont souvent imprécis. De plus, la *jachère* est souvent conçue, d'un point de vue plus "traditionnel" dans le domain de la télédétection, comme un "état" saisonnier de la surface terrestre (c'est-à-dire une classe d'occupation des sols), s'appuyant sur un profil radiométrique très contrasté. L'hypothèse du profil contrasté est peu vérifiée en Afrique de l'Ouest où végétation naturelle et cultures pluviales sont un développement synchronisé avec la saison des pluies.

À la lumière des implications de la pratique de la jachère, ou dans un sens plus large, les terres agricoles non actives (NAAL), sur la sécurité alimentaire et d'autres services environnementaux connexes, on observe un déficit de système de suivi des jachères dans le monde, avec contexte radiométrique particulièrement difficile en Afrique de l'Ouest. Ainsi, l'objectif principal de cette thèse est d'explorer et d'évaluer une stratégie viable pour discriminer les terres agricoles non actives (NAAL) grâce à l'utilisation de séries temporelles d'images à haute résolution spatiale, combinées à des algorithmes de classification d'images d'apprentissage automatique et d'apprentissage profond en Afrique de

l'Ouest.

À cette fin, nous supposons que les récentes missions satellitaires telles que Sentinel-2 de l'ESA, qui fournissent des acquisitions à haute résolution et à haute fréquence dans le domaine optique multispectral sur le globe, ont le potentiel pour résoudre certains de ces problèmes : (i) une résolution spatiale décamétrique, compatible avec la taille moyenne des champs dans la plupart des agrosystèmes d'Afrique de l'Ouest, ce qui permet de faire face aux problèmes de fragmentation du paysage, (ii) un temps de revisite de 5 jours qui, même en tenant compte des problèmes de nébulosité, permet en théorie le suivi de la croissance de la végétation au cours d'une saison culturale, et (iii) une archive de données suffisamment longue (à partir de la fin 2015) pour analyser un cycle de jachère complet.

L'approche générale suivie dans cette thèse se base sur des analyses préliminaires effectuées sur des relevés de terrain et sur des séries temporelles d'images pré-traitées afin d'identifier les caractéristiques radiométriques potentielles pouvant être associées à des terres agricoles non actives (chapitre 3). Ensuite, nous utilisons ces connaissances pour élaborer et tester trois stratégies de classification (chapitres 4, 5 et 6) qui reposent sur différentes manières de concevoir les terres agricoles non actives, sur des méthodes variées en termes de traitement de l'information temporelle contenue dans les séries d'images, et enfin, sur des jeux de données de référence adaptés à chaque méthode.

Chapitre 2

Les analyses menées dans cette thèse se concentrent sur le site d'étude de Koumbia, dans le sud-ouest du Burkina Faso. La région est caractérisée par un paysage fortement hétérogène, composé de petits champs de culture (souvent moins de 1 ha). L'agriculture locale est principalement pluviale, avec le maïs et le coton comme principales cultures.

Dans cette thèse, les données de référence utilisées pour l'entraînement et l'évaluation des modèles de classification d'images, sont issues du jeu de données JECAM-Burkina Faso (Jolivot et al., 2021). Ce jeu de données contient des informations géo-spatialisées sur l'occupation et l'utilisation des sols, y compris les cultures, les terres en jachère et d'autres classes complémentaires de terres non cultivées, pour les années comprises entre 2013 et 2021.

Les donnés radiométriques utilisées sont des séries temporelles d'images Sentinel-2 (SITS) orthorectifiées et avec une interpolation multi-temporelles (gap-filling) de 10 jours, sur la période 2016-2021. Nous avons aussi inclus une série d'indices radiométriques dérivés des bandes originales tels que l'indice de végétation par différence normalisée (NDVI) (voir **Tab. 2.2**).

Nous avons également utilisé les données pluviométriques mensuelles CHIRPS v2.0 et les données de texture du sol ISDA comme variables complémentaires. Ces dernières (ISDA) ont été utilisées exclusivement dans le cadre d'une analyse exploratoire menée au chapitre 3, tandis que les données CHIRPS n'ont été utilisées que dans l'un des essais de classification d'images du chapitre 5.

Pour la classifications des images satellite, nous avons mobilisé le classificateur Random Forest et TempCNN (Pelletier et al., 2019), un algorithme spécialement conçu pour exploiter la dimension temporelle des séries temporelles d'images satellites. Les détails du protocole général sont détaillés dans le **Chapitre 2**, section *General model training protocol*. Les résultats ont été évalués par validation croisée 5-fold.

Chapitre 3

Dans ce chapitre, nous avons exploré l'ensemble de données JECAM et les SITS pour caractériser la pratique de la jachère d'un point de vue radiométrique. Nous avons analysé la manière dont cette pratique est liée aux conditions pédoclimatiques locales et analysé les caractéristiques spectrales, temporelles et spatiales associées afin de mieux comprendre les problèmes de cartographie de cette pratique. En combinant ces deux sources de données, nous avons analysé et testé différentes hypothèses qui pourraient potentiellement être associées à l'utilisation de la jachère, comme celles qui ont été introduites dans le cadre de l'étude bibliographique de la pratique de la jachère dans le chapitre 1.

L'un des premiers obstacles rencontrés dans ce chapitre a été le manque de cohérence spatio-temporelle des données de vérité terrain où, en raison d'une effet combiné du protocole d'échantillonnage utilisé pour la construction de la base de données JECAM et des pratiques agricoles locales, il a fallu traiter les données de référence pour générer un ensemble de données harmonisées dans le temps et dans l'espace pouvant être utilisé pour l'analyse de séries temporelles pluriannuelles (voir **Sec.3.1.2**). Nous avons élaboré une procédure pour estimer la proportion de champs JECAM affectés par un processus de fragmentation/fusion interannuelle qui a montré que cette proportion pouvait atteindre plus de 20% de la surface agricole disponible en fonction de la paire d'années considérée (voir **Sec.3.1.1**).

Dans le cadre de l'analyse spectrale exploratoire, nous avons utilisé des profils d'indice de végétation par différence normalisée (NDVI) comme variables descriptives des jachères. Les analyses préliminaires ont indiqué un faible niveau de separabilité des class d'ocupation des sols lors de la comparaison à l'échelle de l'année (voir **Sec.3.2**). Quand on étend cette analyse à l'échelle pluriannuelle, avec des données de terrain harmonisées, les résultats ont suggéré un effet cumulatif positif, plus perceptible lors des pics saisonniers de NDVI, sur plusieurs années consécutives de mise en jachère. Néanmoins, la grande hétérogénéité des surfaces ne permet pas de délimiter clairement la dynamique de la pratique de la jachère et d'établir un système de discrimination basé sur ces règles.

Enfin, nous avons effectué deux tests complémentaires pour analyser les possibles relations spatiales avec la pratique de la jachère. Plus précisément, nous avons exploré les associations entre la pratique de la jachère et (a) les types de sol, et (b) la position relative de la jachère par rapport aux villages (voir **Sec.3.3**). Aucun de ces tests n'a donné de résultats positifs. Il n'est pas clair si cela est lié à un problème de sous-représentation des champs en jachère dans le jeu des données JECAM et/ou à la taille réduite de la zone d'étude.

Chapitre 4

Dans ce chapitre, nous avons testé une approche basée sur les données en employant des stratégies classiques de classification d'images pour détecter les jachères annuelles, en utilisant des images Sentinel-2 interpolées comme données d'entrée et le jeu de données JECAM comme données de référence. Nous avons testé une stratégie de classification automatique supervisée des images et deux classificateurs d'images différents (Random Forest et TempCNN). Nous avons également testé une méthodologie non supervisée développée par Tong et al. (2020) spécifiquement pour la cartographie des jachères dans la bande du Sahel à l'échelle annuelle.

Dans les deux approches, la jachère (fallow) est considérée comme une classe d'occupation des sols contrastée, au même titre que les classes de cultures (cropped), les forêts à feuilles persistantes (evergreen) ou l'eau (water). Outre les classes déjà mentionnées,

nous avons également inclus comme classes annuelles LCLU les surfaces non-végétalisée (non-vegetated) et la savane herbacée (herbaceous savanna). Tous les tests de classification ont donné des résultats médiocres pour la classe de jachère, avec des scores F1 inférieurs à 0.2 (voir **Fig. 4.3** et **Tab. 4.2**). Dans l'approche supervisée, la précision des cinq autres classes LCLU variait entre 0.6 (non-vegetated) et 0.99 (water). Les modèles résultants ont montré un manque de pouvoir discriminant par rapport à la classe Jachère, ce qui indique que les méthodes de cartographie annuelle *classique* ne sont pas suffisantes pour discriminer cette classe.

Chapitre 5

Dans ce chapitre, nous avons présenté une stratégie de cartographie des terres agricoles non actives (NAAL) basée sur l'analyse des trajectoires. Cette méthodologie répond à certaines des limites liées aux difficultés associées à la cartographie d'une pratique culturale en mettant en œuvre une nomenclature de gestion plus large et en exploitant des données de référence pluriannuelles. Une partie des résultats et les conclusions obtenues ont fait l'objet d'une publication (Castro Alvarado et al., 2023). Dans le manuscrit, les expériences présentées dans la publication ont été étendues à l'algorithme TempCNN comme classificateur d'images alternatif.

La méthodologie a consisté en une cartographie NAAL en deux étapes où nous avons d'abord produit une carte des classes de gestion annuelles (annual management maps) et puis nous avons analysé les changements de gestion interannuels entre les classes afin de détecter les NAAL. La première étape s'appuie sur une nomenclature basée sur 4 classes qui a permis d'obtenir des précisions élevées (score F1 supérieur à 0.7 pour toutes les classes et années considérées) tout en favorisant la séparabilité des classes de gestion, à savoir les classes gérées (managed, M), non gérées (unmanaged, U), à feuilles persistantes (evergreen, E) et non végétalisées (non-vegetated, NV). Cette méthodologie est centrée sur la détection des changements de la classe gérée à la classe non gérée (M→U) où la classe gérée est composée des classes de couverture culturales (crop) de JECAM et la classe non gérée est composée des surfaces en jacheres enregistrées dans la base de données JECAM ainsi que d'autres espaces herbacés naturellement végétalisés qui peuvent facilement être confondus avec des jacheres (voir le flux de travail à la **Fig. 5.1**). La deuxième étape a été réalisée à l'aide d'un système basé sur des règles qui opèrent au niveau du pixel . Lorsqu'un pixel géré (M) précédemment classé devient non géré (U), nous le ré-étiquetons comme NAAL (N).

Étant donné que l'ensemble des données du JECAM n'est pas bien adapté à une analyse pluriannuelle, nous avons pris en charge l'ensemble des données et appliqué différentes hypothèses pour pallier le manque d'harmonisation spatio-temporelle. Ceci a conduit à quatre variations du jeu de données original que nous avons utilisées pour vérifier la précision de notre méthodologie (voir **Fig. 5.3**).

Les scores F1 obtenus pour la classe NAAL varient entre 0,50 et 0,92 en fonction de l'année et de l'ensemble de données de validation utilisé. Nous avons détecté des différences spatiales significatives entre les cartes générées, en fonction de l'outil de classification (Random Forest ou TempCNN) utilisé. Cette différence a été la plus appréciée lors de l'inspection des limites des champs et de la confiance des classes du modèle, où TempCNN s'est démarqué avec une confiance de classe de modèle plus stable dans les zones où aucune donnée de référence n'était disponible. Dans les deux cas, la qualité géospatiale des cartes prédites a permis d'identifier des objets "semblables à des champs" avec de faibles niveaux d'incertitude de classification dans les limites administratives de

Koumbia.

Chapitre 6

Dans ce chapitre, nous testerons deux stratégies différentes pour former des modèles de classification d'images en utilisant plusieurs années consécutives d'images Sentinel-2 (et leurs indices spectraux associés) qui ciblent deux conceptions différentes de terres agricoles non actives. De cette manière, nous avons conçu (i) une approche directe de cartographie des jachères sur plusieurs années où nous utilisons la même typologie en six classes que celle du chapitre 4 en utilisant la base de données JECAM comme données de référence. Dans cette approche, nous avons testé de multiples combinaisons des séries temporelles comme données d'entrée. Dans cette approach directe, nous avons implementé trois stratégie differentes: causale (CA), anti-causale (AC) et combinée (CA/AC). Dans une autre approche pluriannuelle (ii), nous avons élaboré une typologie de trajectoire dans laquelle nous avons identifié deux systèmes agricoles contrastés : les "terres agricoles dynamiques" (Dynamic Agricultural Land, DAL), où des périodes d'inactivité telles que la jachère, la savane herbacée ou la forêt sont observées pour des terres agricoles (cultivées au moins une fois) au cours de la période 2016-2021, et les terres agricoles toujours cultivées en continu (Always cropped) pendant la même période. . Nous avons entraîné un classificateur d'images en utilisant toutes les images Sentinel-2 disponibles pour la période 2016-2021 et nous avons utilisé les données de vérité terrain du JECAM comme données de référence adaptées à cette nomenclature de systèmes de trajectoires. En plus des deux systèmes agricoles sus-mentionnés, nous avons également cartographié les forêts à feuilles persistantes (evergreen), les surfaces non-végétalisées (non-vegetated), la savane herbacée (herbaceous savanna) et l'eau (water). (voir **Fig. 6.2**).

Les résultats obtenus dans le cadre de l'approche de cartographie des jachères pluriannuelles causal/anti-causal ont montré des précisions faibles pour tous les essais considérés, les essais les plus performants atteignant un score F1 de 0.4 (voir **Fig. 6.7**). Néanmoins, en considérant toutes les classes cibles et en regroupant les résultats par essai et par classe (voir **Fig. 6.6**), nous avons observé des différences de performance relativement significatives en fonction de la quantité de données incluses (c.-à-d. le nombre d'années d'entrée de Sentinel-2 SITS), ainsi que de la stratégie employée (c'est-à-dire causal/anti-causal / combiné). Pour le cas spécifique de la classe de jachère, les résultats suggèrent un potentiel avantage avec l'inclusion de plusieurs années de séries temporelles (SITS) en entrée de la procédure d'entrainement.

Dans le cas de l'approche basée sur les systèmes de cultures, le score F1 moyen pour la classe DAL est 0.44, tandis que celui des autres classes varie entre 0.6 et 0.8, à l'exception de la classe "non-végétalisée".

Bien que l'approche fondée sur les systèmes de culture ait permis d'améliorer légèrement les performances pour la classe cible liée à la jachère par rapport aux stratégies de cartographie des jachères directes, elle ne garantit pas un niveau de précision suffisant pour fournir des estimations fiables des terres en jachère.

Chapitre 7

Dans cette thèse, nous avons exploré différentes manières de modéliser la pratique de la jachère avec des données de télédétection, en mettant l'accent sur la dimension temporelle de cette pratique (voir **Fig. 7.1**).

Cela a abouti aux différentes stratégies de cartographie présentées dans les chapitres

4, 5 et 6 où la jachère est considérée comme un type d'occupation du sol défini par des données radiométriques saisonnières (Ch. 4/Ch. 6), défini via la détection de changements interannuels dans leur gestion (NAAL; Ch. 5) ou bien via la détection de systèmes de culture (Ch. 4/Ch. 6).

Nos résultats ont montré que les approches pluriannuelles ont un pouvoir discriminant accru par rapport aux approches de cartographie annuelle des jachères. La stratégie de cartographie des NAAL basée sur les trajectoires (Ch. 5) s'est avérée être la plus performante avec des scores F1 variant entre 0.5 et 0.9 en fonction de l'année et de l'ensemble de données de validation utilisé (voir la description des différents ensembles de données de validation à la **Sec. 5.1.3**). Dans cette thèse, nous fournissons une première preuve que pour une quantification, même approximative, des surfaces de NAAL, la détection régulière des terres cultivées actives par rapport aux autres types des surfaces vegetalisé non gérée peut suffire.

Les difficultés rencontrées lors de l'exploitation des données de référence JECAM, nous permettent d'émettre un certain nombre de recommandations concernant les protocoles d'échantillonnage de la vérité terrain. De plus, nos méthodologies doivent encore être testées dans d'autres régions d'Afrique subsaharienne, où il est probable qu'elles soit confrontées à un manque de données de référence. Pour cette raison, nous considérons que les expériences futures ne devraient pas seulement se concentrer sur des approches supervisées, mais aussi incorporer les adaptations méthodologiques nécessaires à l'élaboration d'une méthodologie non supervisée.