

**Faculté des bioingénieurs**

# **Automatic field extraction using deep learning in the context of smallholder agriculture**

Auteur : Gauthier HEROUFOSSE

Promoteur : Prof. Pierre DEFOURNY

Lecteurs : Prof. Patrick BOGAERT, Dr. Julien RADOUX

Année académique 2022-2023

Mémoire de fin d'études présenté en vue de l'obtention du diplôme de  
Bioingénieur : sciences et technologies de l'environnement.





# Abstract

Sub-Saharan Africa is the center of the world's population growth but is particularly vulnerable to food shortages. In the framework of collaboration between the Institut National de la Statistique du Mali (INSTAT) and UCLouvain, the goal is to achieve a better understanding of agricultural dynamics and improve agricultural statistics. Automatic field extraction from satellite imagery is an essential tool to achieve agricultural landscape mapping, which acts as a decision-support tool to help improve food security. A newly acquired set of reference agricultural field delimitations of  $16.000km^2$  has been prepared by the Malian partner of the Institut d'Economie Rurale (IER) in order to make the first step in that direction. This thesis aimed to evaluate the performance of an automatic field extraction methodology from High Resolution satellite images, in the context of smallholder agriculture, which is characteristic of the sub-Saharan region.

The methodology relies on a Deep Learning model trained with satellite images from 2021 captured by the Planet satellite and the reference dataset. The first step consisted of realizing a calibration analysis to set the best hyperparameters for the model training. The baseline method was then evaluated with carefully selected metrics showing the performance of our model. The last step includes an assessment of the model's portability within Africa, as well as an exploration of possible improvements by adjusting the quality of the input data and the imagery's spatial resolution.

Overall, the results were mixed. The model is mostly capable of finding the fields' extent but struggles to capture field boundaries correctly. The portability of the model is promising, but it suffers from land cover and topography differences. A major drawback is the lack of reliability of the reference dataset. Finally, recommendations and improvement opportunities were discussed for a potential continuation of experiments.

# Acknowledgements

Many people have contributed to the production of this master's thesis. I'd like to start by thanking them in the most appropriate way.

First of all, I would like to thank my supervisor, Pierre Defourny, for his precious advice and recommendations throughout this master's thesis. His patience and confidence helped me to keep the boat on course when it was about to sink. I am also extremely grateful to Quentin Deffense, who helped me with numerous discussions to drive the project forward, and his availability was greatly appreciated.

I'd like to thank my Viennese friends Niels and Jules, who have helped me through this last year with renewed motivation, and with whom I've spent so much wonderful time during the first semester.

Thanks to my close friends from Les Raclures and the NSV team. Our friendship is always a source of entertainment and a ray of light for me, and I hope we'll continue to have great times together in the years to come.

Thank you, Ana, for your much-needed presence over these last few chaotic years. I hope you've found your way.

Thanks to my family for their unconditional love and presence, and for those unforgettable summer holidays.

I hope the readers, Mr Bogaert and Mr Radoux, enjoy reading this thesis as much as I enjoyed writing it.

Computational resources have been provided by the supercomputing facilities of the Université catholique de Louvain (CISM/UCL) and the Consortium des Équipements de Calcul Intensif en Fédération Wallonie Bruxelles (CÉCI) funded by the Fonds de la Recherche Scientifique de Belgique (F.R.S.-FNRS) under convention 2.5020.11 and by the Walloon Region.

# Contents

|                                                                   |             |
|-------------------------------------------------------------------|-------------|
| <b>Contents</b>                                                   | <b>iii</b>  |
| <b>List of Figures</b>                                            | <b>v</b>    |
| <b>List of Tables</b>                                             | <b>vii</b>  |
| <b>Acronyms</b>                                                   | <b>viii</b> |
| <b>1 Introduction</b>                                             | <b>1</b>    |
| <b>2 Literature review</b>                                        | <b>2</b>    |
| 2.1 Instance Segmentation . . . . .                               | 2           |
| 2.2 Semantic segmentation . . . . .                               | 3           |
| 2.2.1 Overall Definition . . . . .                                | 3           |
| 2.2.2 Pixel-based vs object-based classification . . . . .        | 4           |
| 2.3 Classical image segmentation methods . . . . .                | 6           |
| 2.3.1 Edge based segmentation methods . . . . .                   | 7           |
| 2.3.2 Region-based segmentation methods . . . . .                 | 8           |
| 2.4 Deep Learning based image segmentation methods . . . . .      | 8           |
| 2.4.1 AlexNet, VGGNet & GoogLeNet . . . . .                       | 9           |
| 2.4.2 Fully Convolutional Networks . . . . .                      | 9           |
| 2.5 Post-processing methods for image segmentation . . . . .      | 12          |
| 2.5.1 Cutoff method . . . . .                                     | 12          |
| 2.5.2 Watershed transformation . . . . .                          | 12          |
| 2.6 Performance metrics . . . . .                                 | 12          |
| 2.6.1 Evaluation methods for image segmentation . . . . .         | 13          |
| 2.6.2 Supervised Objects metrics . . . . .                        | 13          |
| 2.6.3 Pixel metrics . . . . .                                     | 16          |
| 2.7 Related work . . . . .                                        | 17          |
| <b>3 Objectives</b>                                               | <b>19</b>   |
| 3.1 General context . . . . .                                     | 19          |
| 3.2 Objectives . . . . .                                          | 19          |
| <b>4 Materials and methods</b>                                    | <b>20</b>   |
| 4.1 Materials . . . . .                                           | 20          |
| 4.1.1 Region of interest . . . . .                                | 20          |
| 4.1.2 Reference datasets collection and quality control . . . . . | 23          |
| 4.1.3 Planet Data Pre-processing . . . . .                        | 25          |
| 4.1.4 Additional images . . . . .                                 | 27          |
| 4.2 Methods . . . . .                                             | 28          |
| 4.2.1 Model selection . . . . .                                   | 29          |

|                                                  |           |
|--------------------------------------------------|-----------|
| 4.2.2 Baseline method                            | 31        |
| 4.2.3 Performance evaluation                     | 35        |
| 4.2.4 Additional experiments                     | 36        |
| 4.2.5 Softwares used                             | 37        |
| 4.2.6 Flowchart                                  | 38        |
| <b>5 Results and discussion</b>                  | <b>39</b> |
| 5.1 Baseline method results                      | 39        |
| 5.1.1 Calibration analysis results               | 39        |
| 5.1.2 Assessment of the baseline method          | 44        |
| 5.2 Additional analysis                          | 51        |
| 5.2.1 Quality Test Set assessment                | 51        |
| 5.2.2 Very High Resolution images                | 52        |
| 5.2.3 Model portability within Africa            | 53        |
| 5.3 Discussion and perspectives                  | 55        |
| <b>6 Conclusion</b>                              | <b>58</b> |
| <b>Bibliography</b>                              | <b>59</b> |
| <b>A Additional material and methods figures</b> | <b>70</b> |
| <b>B Results of the baseline method</b>          | <b>73</b> |

# List of Figures

|     |                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |    |
|-----|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 2.1 | Possible ways of realising instance segmentation . . . . .                                                                                                                                                                                                                                                                                                                                                                                                             | 3  |
| 2.2 | Relationship between objects under consideration and spatial resolution: (a) low resolution: sub-pixel approaches required since pixels are much larger than objects. (b) medium resolution: sizes of the pixels and the objects are equal, and pixel-by-pixel approaches are appropriate. (c) high resolution: because pixels are much smaller than objects, they must first be rationalized into groups of pixels and finally objects. From Blaschke (2010). . . . . | 6  |
| 2.3 | Atrous Spatial Pyramid Pooling (ASPP). ASPP uses multiple parallel filters running at various rates to classify the central pixel (orange) to deal with multi-scale features. Different colors are used to indicate the effective Field-Of-Views. Source: Chen, Papandreou, Kokkinos, et al. (2018). . . . .                                                                                                                                                           | 10 |
| 2.4 | U-net architecture. Each blue box corresponds to a multi-channel feature map. The number of channels is denoted on top of the box. The x-y-size is provided at the lower left edge of the box. White boxes represent copied feature maps. The arrows denote the different operations. Source: Ronneberger et al. (2015). . . . .                                                                                                                                       | 11 |
| 2.5 | Classification of evaluation criteria in image segmentation. From Wang et al. (2020). . . . .                                                                                                                                                                                                                                                                                                                                                                          | 13 |
| 4.1 | Location of the ROIs inside the malian territory. . . . .                                                                                                                                                                                                                                                                                                                                                                                                              | 21 |
| 4.2 | Location of the main ROI. The border between the Sikasso and the Koulikoro region is visible in the bottom left corner inside the ROI. . . . .                                                                                                                                                                                                                                                                                                                         | 22 |
| 4.3 | Average monthly precipitation and temperature in Koutiala, from <i>Climate data for cities worldwide - Climate-Data.org</i> (2023). . . . .                                                                                                                                                                                                                                                                                                                            | 22 |
| 4.4 | Illustration of the problem related to the split between sub-perimeters for the data segmentation. The yellow segment lines trace the delimitation between those imaginary squares. . . . .                                                                                                                                                                                                                                                                            | 24 |
| 4.5 | Zoom on the quality controlled reference dataset displaying the field size of each individual polygon. The complete dataset is available in <a href="#">Additional material and methods figures</a> . . . . .                                                                                                                                                                                                                                                          | 25 |
| 4.6 | Zoom on the Quality set. The complete Quality set is available in <a href="#">Additional material and methods figures</a> . . . . .                                                                                                                                                                                                                                                                                                                                    | 26 |
| 4.7 | Illustration of the remaining aggregates present on the Planet satellite images. . . . .                                                                                                                                                                                                                                                                                                                                                                               | 26 |
| 4.8 | Data augmentation example. The left image represent the original, and the right the corresponding rotated and zoomed in image. . . . .                                                                                                                                                                                                                                                                                                                                 | 28 |
| 4.9 | Flowchart of the complete methodology applied throughout this master's thesis. . . . .                                                                                                                                                                                                                                                                                                                                                                                 | 38 |
| 5.1 | Validation metrics comparison obtained with different monthly associations. . . . .                                                                                                                                                                                                                                                                                                                                                                                    | 40 |
| 5.2 | Analysis of the variability of the validation metrics using the same LR for every model. . . . .                                                                                                                                                                                                                                                                                                                                                                       | 41 |
| 5.3 | Influence of the starting LR on the validation metrics. . . . .                                                                                                                                                                                                                                                                                                                                                                                                        | 41 |
| 5.4 | Comparison of prediction maps with and without the HSV layer. . . . .                                                                                                                                                                                                                                                                                                                                                                                                  | 42 |
| 5.5 | Comparison of prediction maps with enhanced recognition of boundaries and extent. . . . .                                                                                                                                                                                                                                                                                                                                                                              | 43 |

|      |                                                                                                                                                                                                                                                            |    |
|------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 5.6  | Validation metrics for the baseline method. The model chosen is the one with the best result in terms of MCC extent, represented by the vertical red line.                                                                                                 | 44 |
| 5.7  | Final result obtained with the baseline method. The extracted fields and fields from the reference layer are displayed according to their area. The rose values on the prediction raster correspond to the extent, and the green/yellow to the boundaries. | 45 |
| 5.8  | Zoom (1) over an area of special interest in the results.                                                                                                                                                                                                  | 46 |
| 5.9  | Zoom (2) over an area of special interest in the results.                                                                                                                                                                                                  | 47 |
| 5.10 | Zoom (3) over an area of special interest in the results.                                                                                                                                                                                                  | 48 |
| 5.11 | Zoom (4) over an area of special interest in the results.                                                                                                                                                                                                  | 49 |
| 5.12 | Prediction raster from the testing of a model trained with VHR Maxar images. The rose values correspond to the extent, and the green/yellow to the boundaries.                                                                                             | 52 |
| 5.13 | Zoom over an area of special interest in the results from the VHR images                                                                                                                                                                                   | 53 |
| 5.14 | Map of the results over the Sikasso region. The extracted fields and fields from the reference are displayed according to their area.                                                                                                                      | 54 |
| 5.15 | Example of data corruption over the Sikasso region. The False Color composite from November shows unusual values.                                                                                                                                          | 55 |
| A.1  | 1-month Precipitation according to CPC (Mar. 21 - Apr. 20, 2021).                                                                                                                                                                                          | 70 |
| A.2  | 1-month Precipitation according to CPC (Oct. 26 - Nov. 25, 2021).                                                                                                                                                                                          | 71 |
| A.3  | Complete quality controlled reference dataset displaying the field size of each individual polygon.                                                                                                                                                        | 71 |
| A.4  | Complete Quality set displaying the field size of each individual polygon.                                                                                                                                                                                 | 72 |
| B.1  | Validation metrics comparison obtained with different monthly associations.                                                                                                                                                                                | 73 |
| B.2  | Analysis of the variability of the validation metrics using the same LR for every model.                                                                                                                                                                   | 74 |
| B.3  | Influence of the starting LR on the validation metrics.                                                                                                                                                                                                    | 74 |
| B.4  | Influence of the starting LR on the validation metrics, zoom around a LR of $10^{-3}$ .                                                                                                                                                                    | 75 |
| B.5  | Scheduler parameters for the weight decays (0.31%). Number of epochs after which the LR decrease.                                                                                                                                                          | 75 |
| B.6  | Comparison of validation metrics with and without the HSV layer.                                                                                                                                                                                           | 76 |
| B.7  | Comparison of validation metrics obtained with enhanced recognition of boundaries and extent.                                                                                                                                                              | 76 |
| B.8  | Reference Quality Test Set                                                                                                                                                                                                                                 | 77 |

# List of Tables

|     |                                                                                             |    |
|-----|---------------------------------------------------------------------------------------------|----|
| 5.1 | Quantitative assessment of the baseline method using pixel-based and object-based metrics.  | 50 |
| 5.2 | Quantitative assessment of the Quality Test set using pixel-based and object-based metrics. | 51 |
| 5.3 | Quantitative assessment of the Sikasso region using pixel-based and object-based metrics.   | 56 |

# Acronyms

**CNN** Convolutional Neural Network

**DL** Deep Learning

**FCN** Fully Convolutional Network

**FPR** False Positive Rate

**GEOBIA** Geographic Object-Based Image Analysis

**HR** High Resolution

**LR** Learning Rate

**MCC** Matthew's Correlation Coefficient

**NDVI** Normalised Difference Vegetation Index

**NIR** Near InfraRed

**OA** Overall Accuracy

**OBIA** Object-Based Image Analysis

**RGB** Red Green Blue

**ROI** Region Of Interest

**RS** Remote Sensing

**SSA** Sub-Saharan Africa

**VHR** Very High Resolution



# Chapter 1: Introduction

Smallholder agriculture is the foundation of food production in Africa. About 90% of the food produced in developing countries comes from this complex agricultural system. In 2016, there were 51 million farms in Africa, 80% of which were less than 2 hectares in size (Lowder et al., 2016). Yet, the number of farms is still increasing while combined with a drop in farm size (Eastwood et al., 2010). Meanwhile, Africa has become the center of the world's demographic growth, and its population could reach 26% of the world's population by 2050 (Abramova, 2022). Sub-Saharan Africa (SSA) is considered to be an area of great concern, as it is particularly vulnerable to food shortages (Economist, 2017). In this context, crop monitoring and mapping systems are essential decision-support tools to help improve food security in developing countries. Agricultural landscape mapping is useful to realize yield predictions and perform long-term planning to make agriculture more resilient.

The production of an object-based map of agricultural land parcels would allow access to various applications: improvement of agricultural statistics, sampling design for yield estimation, cadastral application, and crop insurance, among others. However, accessing such a map is a significant challenge in smallholder agriculture areas for various reasons: small field size, mixed cropping, heterogeneity in management practices, landscape fragmentation, and the widespread presence of trees within the fields.

In recent years, Remote Sensing (RS) has emerged as one of the major tools in the implementation of such systems. Especially, automatic field extraction from satellite imagery is a high-interest task as it would reduce the reliance on time-consuming manual operations. This was previously done in various countries worldwide but is still not achieved in SSA and other smallholder farming areas. With the emergence of High-Resolution (HR) images, new methods are now being developed using Deep Learning (DL) models to accomplish the objective. This master's thesis attempts to develop the best methodology for extracting fields in the context of smallholder agriculture.

The master's thesis starts with a broad literature review about segmentation methods, with a special interest in DL methods. This will be followed by a brief introduction to the context of this thesis and a statement of objectives. Afterwards, the materials and methods will be presented, followed by the results and their associated interpretation, as well as a discussion of prospects for improvement.

# Chapter 2: Literature review

This chapter reviews the state of the art in the field extraction and agricultural segmentation domain. This will allow us to precisely define the thesis' scientific contribution based on what has been previously achieved and develop an appropriate protocol. It begins with the definition of instance segmentation and the methods to achieve it in the RS context. The image segmentation techniques will be detailed, followed by the relevant post-processing techniques. Performance metrics to assess a field extraction result will then be outlined. Finally, related work carried out in recent years will be presented, and the scientific contribution of the thesis will conclude the section.

## 2.1 Instance Segmentation

Instance segmentation is the computer vision task which aims to identify and delineate each individual object within an image. It is one of the most challenging tasks in computer vision as it requires correctly identifying each object, while also precisely segmenting each instance. In the RS context, only a few different object types are found on satellite pictures as the resolution doesn't allow for the precise identification of every element. Typically, buildings, water bodies, fields, forests, and other land cover types are common elements that can be discriminated. Instance segmentation is commonly used, for example, to realize field extraction and obtain a complete agricultural parceling to improve agriculture monitoring (Yan et al., 2014; Waldner et al., 2020; Zhong et al., 2023). Field extraction can be either realized by a state-of-the-art Mask R-CNN model (He et al., 2018) with a one-step procedure or with a two-step procedure starting with semantic segmentation followed by post-processing of the result. Figure 2.1 schematically illustrates the ways through which instance segmentation can be achieved.

A distinction needs to be made between object-detection methods and semantic segmentation, as they can sometimes be confused. Object-detection algorithms aim to split the image into distinct real-life objects. Thus, two neighboring pixels of the same class could belong to two different instances or regions. For instance, when there are two merged cars in an image, object-detection methods need to distinguish between them as separate entities. On the other hand, in the semantic segmentation domain, those two merged cars would simply be part of the same class. However, object-detection algorithms are rare in agricultural monitoring applications (Li

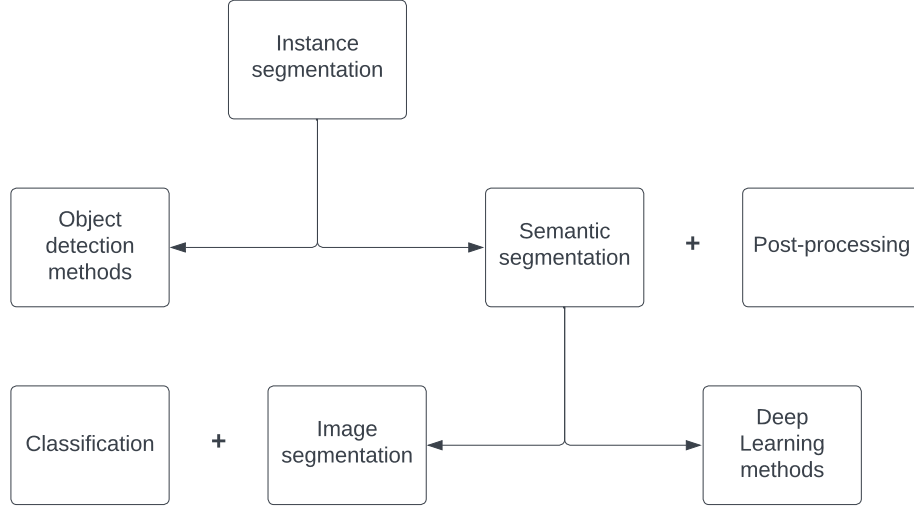


Figure 2.1: Possible ways of realising instance segmentation

et al., 2020). Crops are typically separated by field boundaries, which makes semantic segmentation a relevant method to use because the distinctions between the fields are already visible in the image. Instead, an appropriate post-processing method is employed to obtain instance segmentation, using the boundaries as frontiers between fields. The following sections describe the methods used to achieve instance segmentation using semantic segmentation.

## 2.2 Semantic segmentation

This subsection introduces the concept of semantic segmentation and highlights the advantages of object-based classification over pixel-based classification.

### 2.2.1 Overall Definition

Image segmentation can be defined as a process of partitioning the image into some non-intersecting regions such that each region is homogeneous and the union of no two adjacent regions is homogeneous. More formally, Pal et al. (1993) and Spirkovska, 1993 explained:

"If  $F$  is the set of all pixels, and  $P()$  a uniformity (homogeneity) predicate defined on groups of connected pixels, then segmentation is a partitioning of the set  $F$  into a set of connected subsets of regions  $(S_1, S_2, \dots, S_n)$  such that

$$\bigcup_{i=1}^n S_i = F \text{ with } S_i \cap S_j = \emptyset, \forall i \neq j$$

The uniformity predicate  $P(S_i) = \text{true}$  for all regions  $(S_i)$  and  $P(S_i \cap S_j) = \text{false}$ , when  $S_i$  is adjacent to  $S_j$ ."

Semantic segmentation is therefore informally defined as the task of labeling each pixel of a segmented image. This part of computer vision has recently become an essential method in various scientific fields, being used for a variety of tasks such as medical image diagnostics, automatic mapping of landslides, or pedestrian detection (Pham et al., 2000, Qi et al., 2020). In the remote sensing field, applications of modern semantic segmentation include environmental monitoring, crop cover and type analysis, tree species identification, and building classification in urban spaces. More details can be found in reviews by Dey et al. (2010), Guo et al. (2018), and Yuan et al. (2021).

### 2.2.2 Pixel-based vs object-based classification

The two main classification techniques that emerged since the late 1980's are the pixel-based image analysis and the object-based image analysis. Pixel-based classification uses multispectral information to label a pixel to a class, based on spectral similarities between the pixels from the same class. In the meantime, various classifiers have emerged as state-of-the-art methods in classification. One can cite models like Support Vector Machine (SVM), K-Nearest-Neighbors (KNN), Maximum Likelihood Classifier (ML), Random Forest (RF), and Neural Networks (NN) (Khatami et al., 2016).

The major historical drawback to those techniques is that they do not use contextual and textural information. As a result, pixel-based classification suffers from the so-called 'salt-and-pepper' effect, caused by not taking information from the neighboring pixels. In addition to lowering the global accuracy of the map, it can lead to a significant decrease in the appreciation quality of the classification. For example, a regular land cover classification user could easily recognize himself in a situation where he is confronted with a pixel of urban area appearing in the middle of the Amazon forest. Taking into account the surrounding areas, such a situation would not occur. However, some techniques used to improve classification accuracy have been developed over the years, gradually improving the results obtained with the pixel-based classification methods. Thus, a majority filter can be applied to reduce the 'salt-and-pepper' effect. Texture, slope, and elevation data can be considered to add extra spatial information. A survey of image classification methods and techniques to improve classification performance has been conducted by Lu et al. (2007).

With the advent of metric Very High Resolution (VHR) imagery, Object-based classification was required to classify the image into meaningful objects. In the field of remote sensing (RS), this approach is part of the field known as *Geographic Object-Based Image Analysis* (GEOBIA), distinguishing it from *Object-Based Image Analysis* (OBIA) found in fields such as Computer Vision and Biomedical Engineering (Blaschke, 2010). Further information about

this semantic separation can be found in Hay et al. (2008). The GEOBIA field aims to replicate visual interpretation of RS images through automated or semi-automatic methods. As a result, more accurate and reproducible information is produced, subjectivity is reduced, and there are decreases in labor and time expenditures. Object-based classification is currently reliant on the outcome of an image segmentation algorithm. Pixels are grouped into spectrally homogeneous image regions called Segments. These so-called 'object candidates' are then subjected to further processing steps to be transformed into meaningful objects and subsequently classified. The processing steps encompass spatial, textural, and contextual features, serving as supplementary observations to the spectral signature. Ultimately, since image-objects reflect "meaningful" geographic entities or occurrences at various scales, they become the fundamental analytical units.

Blaschke (2010) demonstrated that there are two major trends contributing to the increase in OBIA-related publications: First, the availability of VHR satellite data. The launch of satellites like IKONOS and QuickBird has resulted in numerous publications since 2001/2002. VHR images enable objects of interest to be significantly larger than the pixel size. As illustrated in Blaschke (2010) (Figure 2.2), only situation (c) allows for clear object identification; the other two situations present insufficient pixel sizes to identify, classify, or characterize observed objects. One can categorize this as a HR situation. Conversely, situations (a) and (b) are considered low-resolution images, where the object is smaller than the resolution cells (Blaschke et al., 2014). According to the Shannon sampling theorem cited by Blaschke (2010), to ensure that an object will be entirely independent of its random position and orientation relative to the sampling scheme, it should be approximately one-tenth of the pixel's size. However, it is challenging to find other sources that corroborate this formulation. More recently, VHR airborne cameras and LiDAR data have reinforced this trend, and GEOBIA is now primarily applied to all VHR and sub-metric images. Objects are then composed of many pixels, making visual validation feasible. Satellites ranging from WorldView-4 with a spatial resolution of 30cm to Planet with a resolution of 3.7m are considered part of the VHR domain covered by GEOBIA.

The popular object-based computer software package called 'eCognition' was released in 2000. This object-based environment analyzes images using semantic information to interpret them correctly. Neighboring pixels are grouped into meaningful objects using color, texture, shape, and/or size (Geospatial, 2022). Due to its success, other commercial software developments were realized afterward, such as Feature Analyst in 2002 (Opitz et al., 2008) and SAGA in 2004 (Boehner et al., 2006). By providing easy access to a large number of researchers, these software packages contributed significantly to the growth of object-based classification techniques.

In the agricultural context, many studies have been conducted to compare the accuracy of pixel-based and object-based methods. For instance, Castillejo-González et al. (2009) used images from the Quickbird satellite in the south of Spain to discriminate among 10 crops. The

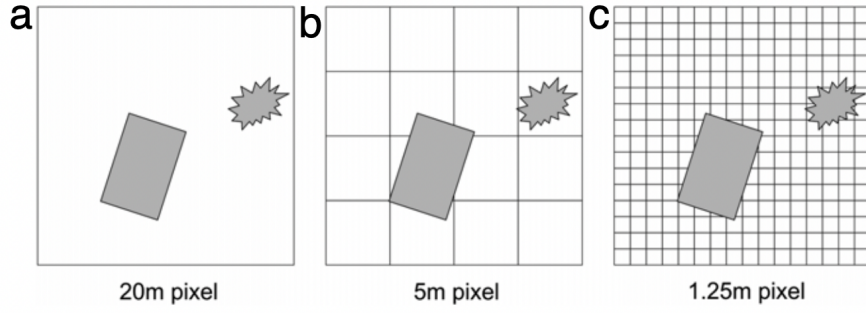


Figure 2.2: Relationship between objects under consideration and spatial resolution: (a) low resolution: sub-pixel approaches required since pixels are much larger than objects. (b) medium resolution: sizes of the pixels and the objects are equal, and pixel-by-pixel approaches are appropriate. (c) high resolution: because pixels are much smaller than objects, they must first be rationalized into groups of pixels and finally objects. From Blaschke (2010).

spatial resolution was 0.7 m, and four initial spectral bands were available. Results showed that object-based classification provides systematically better Overall Accuracy (OA) and Kappa coefficient than pixel-based classification, even with the use of a 5x5 majority filter to reduce the 'salt-and-pepper' effect. Depending on the chosen classifier, it starts from an OA 25% higher with the parallelepiped classifier, to a much smaller difference (<2%) with the Maximum-Likelihood classifier. Castillejo-González et al. (2009) suggests that VHR imagery increases intraclass spectral variability and results in a reduction in classification performance using pixel-based classification methods. Indeed, Duro et al. (2012) used HR images from SPOT-5 to classify broad land cover classes over agricultural landscapes. They showed that based exclusively on OA, there is no advantage to using object-based analysis over pixel-based analysis with decametric resolution, as the results were not statistically different. More recently, Belgiu et al. (2018) used the Time-Weighted Dynamic Time Warping Algorithm (TWDTW) in three different regions across the world to differentiate crop types. Using Sentinel-2 time series for cropland mapping, the algorithm was tried with both object- and pixel-based classification. They report that OB-TWDTW outperformed PB-TWDTW in all test areas in terms of the quality of the classification outputs, as measured by the overall accuracy metric, and in terms of computational time. More generally, it is recognized that object-based classification methods provide more accurate results and more attractive classification maps in the context of VHR image processing.

## 2.3 Classical image segmentation methods

To perform GEOBIA semantic segmentation, one can use the traditional two-step scheme, consisting of image segmentation followed by an object classification using the segments resulting from the segmentation. Simple object classification won't be covered here as it falls outside of the scope of this study. Alternatively, one-step semantic segmentation can be performed

using a DL algorithm (see Section 2.4). This section explores the classical image segmentation techniques that are still used in the RS context.

### 2.3.1 Edge based segmentation methods

The goal of edge-based segmentation methods is to identify edges between homogeneous regions and define segments as the areas within these edges. Edges are considered boundaries between image objects and are situated where abrupt changes in the image values occur. Since edges are local features, they are determined based on local information (Pal et al., 1993). Edge detection methods were categorized by Davis (1975) into two groups: sequential and parallel processing. In a sequential approach, the decision of whether a pixel is an edge or not is influenced by the previous decisions made by the detector. Conversely, in a parallel approach, the decision is solely based on the central pixel and some of its neighboring points. Various types of elementary parallel operators exist, such as the Roberts, Sobel, and Prewitt gradient, as well as the Laplacian operator. Canny (1986) subsequently developed a mathematical approach to edge detection based on three criteria aimed at optimizing edge detection. The first criterion is a low error rate. Detecting all edges in images while minimizing erroneous responses serves as significant performance indicators in the field of edge detection. The number of false positives and false negatives should be minimized. The second criterion involves accurately locating the edge on the image. In other words, the distance between the detected edge's location and the "true center" of the edge should be minimized. The final criterion is the requirement for a single response per edge on the image. For a comprehensive survey and comparison of existing edge detectors, readers are directed to the paper authored by Heath et al. (1997).

In most circumstances, the Canny detection algorithm gives superior results among the classical edge detection techniques (Nguyen et al., 2000, Pellegrino et al., 2004). Essentially, this process is composed of these key steps:

1. Smooth the image with an appropriate Gaussian filter to reduce undesired image details.
2. Determine gradient magnitude and gradient direction at each pixel.
3. If the gradient magnitude at a pixel is larger than those at its two neighbors in the gradient direction, mark the pixel as an edge. Otherwise, mark the pixel as the background.
4. Remove the weak edges by hysteresis thresholding.

Although the Canny edge detector is the most rigorously defined operator and is widely used, some typical problems remain. For instance, the removal of weak edges through hysteresis thresholding relies on fixed values. In images with rich edge information, the adaptability of



the Canny edge detector is less than ideal, potentially leading to the omission of crucial information. Therefore, additional adjustments and optimization methods have been incorporated into the original Canny edge detector to enhance its detection outcomes (Ding et al., 2001; Rong et al., 2014). Various applications of edge-based segmentation can be found in the literature. For instance, Dai et al. (2019) employed the Canny edge detector to identify terrace elevations in the Loess Plateau in China. However, it is primarily employed as a preprocessing step in hybrid methods for object detection. The resulting edge matrix can be used as a foundational element for more advanced object construction techniques, such as region-based algorithms or graph models.

### 2.3.2 Region-based segmentation methods

Region-based segmentation algorithms use another way to regroup pixels with similar properties into meaningful objects. Instead of looking for abrupt changes in the image that characterize edges, they look for shared traits among neighboring pixels. They can be mainly separated into those two categories: region growing and region merging. Region growing techniques (also called bottom-up approaches) start from an initial set of seed pixels, and recursively aggregate the surrounding pixels using a similarity criterion, until a target threshold is reached (Kotaridis et al., 2021). Among region-growing algorithms, Mean Shift is a powerful procedure used to locate image modes corresponding to homogeneous region centers. The mode centers are detected by iterative weighted mean estimation. Two bandwidth parameters  $h_s, h_r$  for spatial and spectral kernels are then used to identify the pixel in the basin of attraction of the modes (Demirović, 2019). On the other hand, region merging and splitting proceed to first divide the image into subregions where the similarity criterion fails. Then, the resulting objects are merged into a set of approximated objects under a merging criterion.

## 2.4 Deep Learning based image segmentation methods

The advances in DL algorithms acted as a paradigm shift in the image processing research field. Particularly, the development of Convolutional Neural Networks (CNNs) and Fully Convolutional Networks (FCNs) rapidly became state-of-the-art methods in semantic segmentation of remote sensing (RS) images, yielding noteworthy results across several applications.

DL methods are a set of representative learning techniques involving multiple processing stages of representation (LeCun et al., 2015). They essentially consist of a multi-layer organization that processes input data into higher and more abstract levels. At each level, relevant features are extracted using a non-linear module and then fed into the subsequent layer. This arrangement is commonly referred to as a deep neural network as it emulates the organization



of our neural system, characterized by numerous layers. Each layer is connected to the next one through an activation function  $a$  and a set of parameters  $W_i$ . Very complex behaviors can be learned with the coordination of a sufficient number of these transformation layers. A key advantage of DL is its independence from human engineering during model training (LeCun et al., [2015]). Consequently, it can readily exploit increases in available computation and data.

In the context of satellite RS imagery, several recent papers have focused on DL methods applied to semantic segmentation (Neupane et al., [2021]; Yuan et al., [2021]; Wang et al., [2022]). We will examine some of the presented model architectures and explain the fundamental ideas behind them.

### 2.4.1 AlexNet, VGGNet & GoogLeNet

AlexNet, VGGNet and GoogLeNet were some of the first CNN's used in image classification. Their structure is based on a series of convolutional layer and fully-connected layers, with pooling operations in between. The fundamentals behind those technical terms will be explained in the section [4.1]. AlexNet (Krizhevsky et al., [2017]) brought some major contributions in the evolution of CNNs :

- use of the Rectified Linear Unit (ReLU) activation function, computationally efficient ;
- use of overlapping max pooling, using a stride smaller than the filter size ;
- introduction of the dropout technique in the fully-connected layers to reduce the overfitting.

VGGNet (Simonyan et al., [2015]) was developed three years later, largely based on AlexNet's structure. VGGNet favours smaller filters to reduce the number of weights needed to train the model, and hence the training complexity. GoogLeNet (Szegedy et al., [2015]) differs from the two mentioned models with the introduction of the inception module. Basically, it applies three filters of different size to the same input, and concatenate the result using a max-pooling operation. The original GoogLeNet model has nine inception modules and three regular convolutional layers. It is therefore already way deeper than the two previous models.

Those three models are worth mentioning as they are at the basis of the more complex FCNs largely used in the semantic segmentation domain.

### 2.4.2 Fully Convolutional Networks

Long et al. ([2015]) introduced the FCN for semantic segmentation purpose. As the name suggest, the FCN differs from the CNN by the presence of only convolutional layers and pooling layer (Sultana et al., [2020]). A general structure for FCN models consists of two components, an encoder and a decoder. Encoders are directly similar to CNNs and extract feature maps. These features are then converted by decoders into dense label maps with the same dimensions

as the input image. Where the encoding part is relatively constant, different methods emerged in order to realize the decoding part. Some of the state-of-the-art models will be introduced together with their key features and advantages.

Chen et al. (2016) and their DeepLab model was one of the first introduced FCN. The encoding part is based on the original VGG16 model (Simonyan et al., 2015) with the addition of 'atrous' convolution to capture a larger field of view and hence providing a better supplied context to the network. Atrous convolutions are kernel filters where spaces are left in between sampled pixels. The width of the filter is defined by the *rate*. To handle both small and large objects, they used a pyramid pooling method, which consists to extract the features using various spatial rate in the atrous convolution layers. The features extracted by each filter are then fused to generate the final result. The proposed "atrous spatial pyramid pooling" (DeepLab-ASPP) is illustrated in Figure 2.3. The decoding part is then made of a fully connected pairwise Conditional Random Field (CRF), described by Krähenbühl et al. (2011).

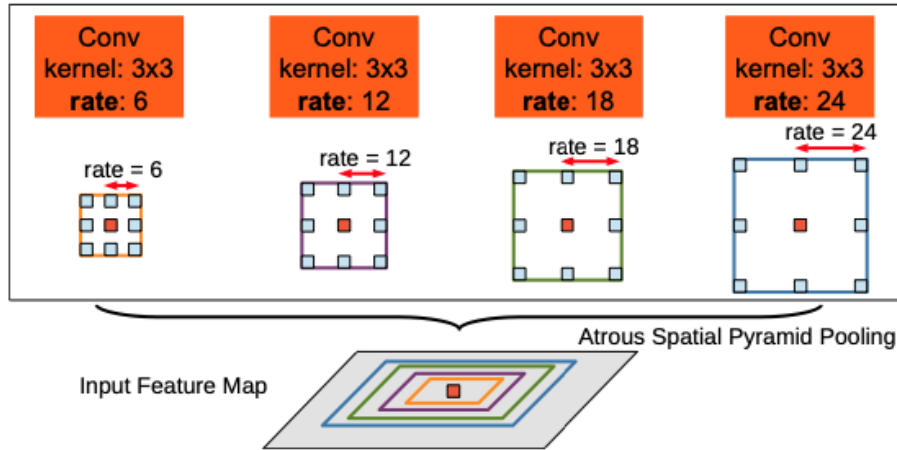


Figure 2.3: Atrous Spatial Pyramid Pooling (ASPP). ASPP uses multiple parallel filters running at various rates to classify the central pixel (orange) to deal with multi-scale features. Different colors are used to indicate the effective Field-Of-Views. Source: Chen, Papandreou, Kokkinos, et al. (2018).

The UNet model is a U-shaped semantic segmentation network with a contracting path and an expansive path (Ronneberger et al., 2015). The contracting path extracts feature maps and condenses spatial information as a traditional CNN. The contracted feature map is then the input for the expansive path, which applies an up-convolution. It consists of an up-sampling of the feature map following by 2x2 up-convolution that halves the number of feature channels. The result is then concatenated with the correspondingly cropped feature map from the contracting path, and two 3x3 convolutions followed by a ReLU are applied (Ronneberger et al., 2015). The global architecture is illustrated in Figure 2.4. Those concepts will be explained

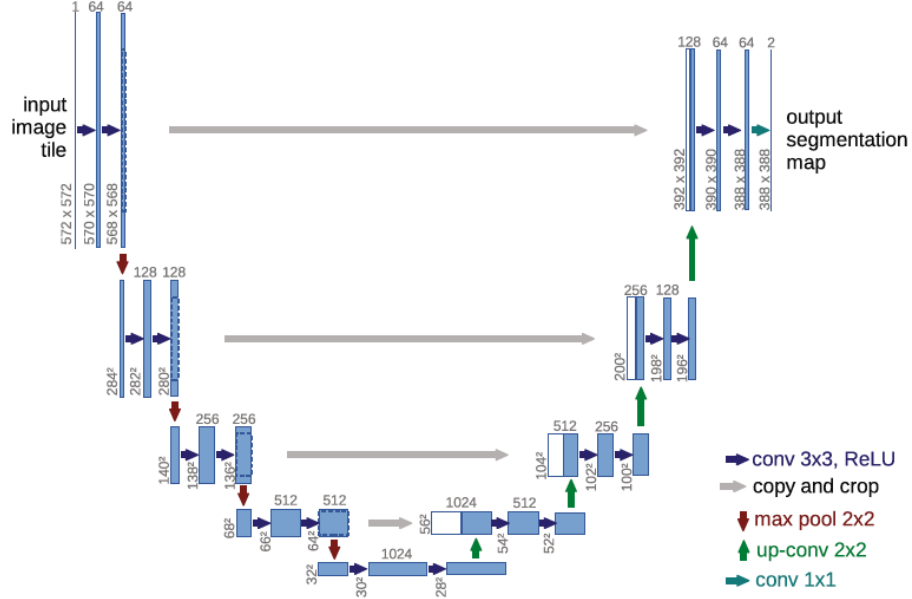


Figure 2.4: U-net architecture. Each blue box corresponds to a multi-channel feature map. The number of channels is denoted on top of the box. The x-y-size is provided at the lower left edge of the box. White boxes represent copied feature maps. The arrows denote the different operations. Source: Ronneberger et al. (2015).

more in details in the section 4.1

SegNet (Badrinarayanan et al., 2017) is also made of a encoder-decoder architecture. The encoder is based on the VGG16 CNN and extracts feature maps and condenses the spatial information. In each layer, a convolutional operation is performed, followed by a max-pooling operation using a 2x2 filter with stride 2, i.e. the filter moves by 2 pixels between consecutive operations. In addition, the element index (i.e., position of an element within the filter window) from the max-pooling operations are stored and used in the up-sampling process of the decoder network. The decoder network follows a symmetric structure to the encoder. As the output should be of same size as the input image, the decoder uses the index to recover spatial information and improve the boundary accuracy. For instance, a 2x2 low-resolution feature map is scaled up to a 4x4 matrix. The elements of the 2x2 map are placed to the position where they were previously pooled from the 4x4 matrix in the corresponding encoding layer.

Many other Deep Learning models exist in the semantic segmentation domain, but it would be laborious and inappropriate to present all of them in this literature review. For more details and specificities of all the models, the reader could consult Ma et al. (2019), Hao et al. (2020), Sultana et al. (2020), Yuan et al. (2021), Neupane et al. (2021), Wang et al. (2022) and Mo et al. (2022).

## 2.5 Post-processing methods for image segmentation

As the neural networks might produce discontinuous boundaries, post-processing methods are necessary to generate closed boundaries and isolate individual fields. Those methods exploit the semantic segmentation outputs to achieve instance segmentation.

### 2.5.1 Cutoff method

Waldner et al. (2020) employed the *cutoff* method, which involves using thresholding techniques on probability maps for fields and boundaries layers (extent mask and boundary mask) in an image. This approach aimed to separate and identify individual fields within the segmented image. By applying a threshold to the extent map, the method was able to detect field boundaries and differentiate them from non-field objects. Moreover, a threshold was also applied to the boundary mask to enhance the definition of boundaries between neighboring fields. The method then computed the symmetric difference between these two binary layers, effectively identifying pixels present in either of the layers but not in their intersection. As a result, this technique produced distinct fields with well-defined, closed boundaries.

### 2.5.2 Watershed transformation

The watershed transformation is a powerful algorithm that is also related to a region-based segmentation method. Applied on a greyscale image, it seeks for pixel and regional similarities. The gradient image is usually used in the watershed transformation because the main criterion for the segmentation is the homogeneity of the gray values of the objects present in the image. The image is viewed as a topographic map that is flooded by water from its minima. A dam is then built between any two basins that join as long as the water continues to flood. When the procedure is complete, i.e. when the entire topographic map is filled in water, each basin represents an image object. However, as the image is typically noisy, the algorithm is likely to suffer from over-segmentation. Some alternatives were developed to tackle this over-segmentation problem, including marker based methods, topological distance, shortest path algorithms or spanning trees, described in Meyer (2012). As this algorithm has been chosen as our post-processing technique, it will be explained in details in section 4.2.

## 2.6 Performance metrics

As more and more segmentation algorithms are being proposed, it is important to discuss the assessment methods to fix the scope and be able to compare different models between each other. However, there is presently no fixed consensus about evaluation methods in the image segmentation field. This section will hence try to review only the most used methods and

metrics in the semantic segmentation domain. For more extensive studies on the topic, one can consult Chen, Ming, et al. (2018), Costa et al. (2018) and Jozdani et al. (2020).

### 2.6.1 Evaluation methods for image segmentation

Zhang (1996) is one of the first to submit a survey on evaluation methods for image segmentation, and Chen, Ming, et al. (2018) as well as Wang et al. (2020) used it as a support to publish their own conception of the evaluation methods. They all present a similar hierarchy of evaluation criteria (Figure 2.5).

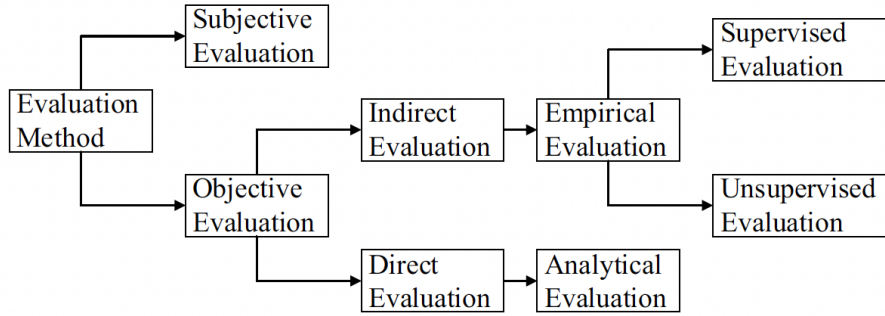


Figure 2.5: Classification of evaluation criteria in image segmentation. From Wang et al. (2020).

The subjective evaluation is, as its name suggests, a qualitative evaluation where the segmentation results are appreciated by a human eye. The main disadvantage of this approach is that visual assessment is nonobjective by nature. The evaluation made by a human evaluator to another will notably vary as each individual has its own unique criteria to assess the quality of the segmented image. Therefore, this method requires a large number of reviewers to be representative of the population, and ends up to be a laborious and time consuming process.

In the RS field, direct evaluation is not really used, as it is the evaluation of the performance of the segmentation algorithm, and not of its results. As the computing time and spatial complexity could be relevant to discriminate two models presenting extremely close results, it is definitely not the first highlighted argument. The indirect evaluation on the other hand judges an algorithm by another mean. The performance of the segmentation result is assessed by calculation or experiment, such as evaluation metrics. This quantitative evaluation can be divided into supervised and unsupervised assessment, on whether a reference dataset is needed or not. Only the supervised methods will be developed here, as we plan to use a reference for evaluation purpose.

### 2.6.2 Supervised Objects metrics

The process of supervised evaluation is divided into 3 steps: establishment of a reference dataset, object matching, and discrepancy calculation (Chen, Ming, et al., 2018). The reference dataset  $R$  for a segmentation purpose consists of a sufficient amount  $n$  of vector polygons

objects, and can be expressed as  $R = \{r_i; i = 1, \dots, n\}$ . It is usually obtained through visual interpretation of remotely sensed images in a similar region than the used one for the training purpose, or digitalized from field data. Because it is subject to manual manipulation, it is a subjective and time-consuming task. Once the reference dataset is fixed, it is right to extract the overlapping object set,  $S_i = \{s_j; \text{area}(s_j \cap r_j) \neq 0\}$ , which regroups a spatial overlapping relationship between each reference object and its corresponding segmented object(s). Ideally, each reference object should be referred to one and only one corresponding object, but in most case, a large number of objects will have multiple overlapping instances, and it is therefore necessary to set a selection process to obtain the best corresponding object set  $S_{i \text{ Max}}^* = \{s_j; \max(\text{area}(s_j \cap r_j)), s_j \in S_i\}$  (Chen, Ming, et al., 2018). The Maximum Overlapping Area Method is a standard method developed by Lucieer et al. (2002), which link the overlapping object that has the largest overlapping area with the reference object. This method is widely used for its simple principle and as it matches only one object to each reference object, while requiring low computation complexity. Chen, Ming, et al. (2018) exposes a few more Object Matching methods that won't be covered here.

Next, the selection of the discrepancy index is subject to a lot of debates, and there is currently no consensus. To simplify, the supervised methods can be divided into geometric and non-geometric methods (Costa et al., 2018). Geometric methods rely on quantitative metrics that outline geometric correspondence between detected objects and referenced polygons. The position, shape and area are compared and there should be ideally no difference among them. Those metrics are usually based either on overlapping area and/or the distance between their centroids. Non-geometric methods don't require an overlay operation between a polygon from reference data and an object resulting from the segmentation. They either focus on the content of the objects, such as representing the segmented object in the spectral space or use a classifier to classify the segmented objects. Some of the key geometric indices will be presented here after.

The Area Fit Index (*AFI*) defined by Lucieer et al. (2002) is based on the Maximum Overlapping Area Object Matching method presented earlier. It assesses the segmentation result by comparing the area of the reference and its corresponding object :

$$AFI_i = \frac{\text{area}(r_i) - \text{area}(s_j)}{\text{area}(r_i)}, s_j \in S_{i \text{ Max}}^*$$

The value of the *AFI* index for the ideal segmentation is 0. If the  $AFI_i$  is positive, it indicates that the reference object is over segmented, while a negative value express undersegmentation. The global *AFI* index corresponds to absolute the mean of all  $AFI_i$  indices.

Zhang et al. (2015) introduced two metrics by adapting the *Precision* and *Recall* measures. *Precision* measures the quality of the segments generated (part of the generated segment that is rightly in the reference object) while the *Recall* measures how well the reference object is represented by the segment, such as :

$$Precision_{ij} = \frac{area(r_i \cap s_j)}{area(s_j)}, s_j \in S_{i\ Max}^*$$

$$Recall_{ij} = \frac{area(r_i \cap s_j)}{area(r_i)}, r_i \in R_i$$

Then, they combined those two indices using the weighted mean by area of all  $Precision_{ij}$  and  $Recall_{ij}$  to obtain an index to express the quality of the segmentation,  $F-measure_{PR}$ .  $\alpha$  is set to 0.5 to give the same weight to both  $Precision$  and  $Recall$ . The closer from 1  $F-measure_{PR}$  is, the optimal the segmentation result is.

$$F-measure_{PR} = \frac{1}{\alpha \frac{1}{Precision} + (1 - \alpha) \frac{1}{Recall}}$$

The Quality Rate ( $QR$ ) index, defined by Weidner (2008), compares the overlap area to the union area of the reference and the corresponding object. The bigger the union is in compare to the intersection, the worst the index is. It ranges from  $[0, 1]$  and 0 is the ideal value, when union = intersection. To obtain the global index, compute the absolute mean of all  $QR_{ij}$  calculated as follow :

$$QR_{ij} = 1 - \frac{area(r_i \cap s_j)}{area(r_i \cup s_j)}, s_j \in S_{i\ Max}^*$$

Yang et al. (2015) introduced the Segmentation Evaluation Index ( $SEI$ ) defined by :

$$SEI_{ij} = \begin{cases} \sqrt{\left(1 - \frac{area(r_i \cap s_j)}{area(r_i)}\right)^2 + \left(1 - \frac{area(r_i \cap s_j)}{area(s_j)}\right)^2}, & s_j \in Sc_i \cap Sd_i \\ 1 & , s_j \notin Sc_i \cap Sd_i \end{cases}$$

with  $Sc_i = \{s_j : area(r_i \cap s_j)/area(s_j) > 0.5\}$ ,  $Sd_i = \{s_j : area(r_i \cap s_j)/area(r_i) > 0.5\}$ . This ensures that the corresponding object shares 50% of its surface with the reference object, while the reference object follows the same rule. The global  $SEI$  index is then the mean of the  $SEI_{ij}$ , and the segmentation is considered as best when  $SEI$  tends to 0.

Those metrics are all very restrictive concerning the over- and the under-segmentation. Following the same approach, Pavlova et al. (2021) introduced the  $RecRate$  index, which consists to count the number of correctly matched reference and predicted objects, with a one-to-one correspondence. To do so, they used the Jaccard coefficient and considered that two field  $f_{ref}^i$  and  $f_{pred}^i$  match if  $J(f_{ref}^i, f_{pred}^i) > 0.5$ , with :

$$J(a, b) = \frac{S(a \cap b)}{S(a \cup b)}$$

Under the condition  $J(a, b) > 0.5$ , the set  $F_{ref} \leftrightarrow F_{exp}$  can be designed as one-to-one correspondence if the fields from  $F_{exp}$  do not overlap. The  $RecRate$  is then defined as :



$$RecRate = 2 \frac{N_{one2one}}{N_{ref} + N_{exp}} * 100\%$$

with the set  $N_{one2one}$  expressing the number of correspondences in the One-to-one set. However, they also introduced a second metric, the  $RecRate_{soft}$ , which ignores the errors due to over- and under-segmentation. The following criterion is defined :

$$\Omega(a, b) = \max\left\{\frac{area(r_i \cap s_j)}{area(r_i)}, \frac{area(r_i \cap s_j)}{area(s_j)}\right\}$$

The field  $f_{ref}^i$  corresponds to the field  $f_{pred}^i = \operatorname{argmax}(\Omega(f_{ref}^i, f))$ , under the condition  $\Omega(f_{ref}^i, f_{pred}^i) \geq 0.5$ . Similarly, the field  $f_{pred}^i$  following the same rule. The sets  $F_{ref}$  and  $F_{pred}$  are so computed, and it allows multiple correspondences. This correspondence set is denoted as  $F_\Omega$ , and the number of  $ref$  fields from this set by  $N_{ref}^\Omega$ . The  $RecRate_{soft}$  is then defined as :

$$RecRate_{soft} = \frac{N_{ref}^\Omega}{N_{ref}} * 100\%$$

However, this set does not take the False Positives into account, i.e., fields from the prediction dataset  $S_i$  that do not share 50% of area with any field from the reference dataset. Those are denoted as  $F_{FP} = S_i \setminus F_\Omega$ . The False Positive Rate ( $FPR$ ) metric accounts for those missing fields :

$$FPR = \frac{N_{FP}}{N_{ref}} * 100\%$$

with  $N_{FP}$  being the number of fields in the  $F_\Omega$  set. In the context of a imperfect reference dataset, this metric could be really useful to understand how well the testing area is covered.

An attentive reader could easily observe similarities between the indices, and understand why there is actually no consensus for one or the other index. It is partly due to the fact that it is still arguable what is known as an "optimal" segmentation (Jozdani et al., 2020). Because land objects have varied structures and properties, optimal segmentation cannot be generated universally. Each land objects has therefore its own interpretation of "optimally segmented". The term "optimal" in the context of segmentation is therefore very application-dependent and should be carefully defined for each application considering the needs and requirements of the research project.

### 2.6.3 Pixel metrics

Along with object metrics, it is quite common to evaluate the segmentation result using a rasterised version of the result, as the pixel metrics are broader recognised than object metrics in the assessment process. Most of the pixel metrics are derived from the confusion matrix, resulting from the comparison of the rasterised result and the corresponding rasterised reference



dataset. For each class, True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN) are counted and then used to compute performance metrics. The simplest methods are probably the *OA*, the *Precision* and the *Recall*, following those respective equations :

$$OA = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

*OA* corresponds to the total number of right predictions divided by the total number of predictions. *Precision* describes the number of correctly predicted pixels divided by the total number of pixel predicted within the class, and *Recall* describes the number of correctly predicted pixels divided by the number of pixels truly belonging to the class. Those three metrics are between 0 and 1, 1 expressing the perfect prediction result. The F1-score (or DICE coefficient) and the Intersection over Union (IoU) are two similar metrics computed as so :

$$F1 - score = \frac{2 * TP}{TP + FP + FN}$$

$$IoU = \frac{TP}{TP + FP + FN}$$

Both metrics are also between 0 and 1 and the F1-score represents the harmonic mean of Precision and Recall. Lastly, Matthew's Correlation Coefficient (MCC) is the most robust pixel metric in the binary<sup>1</sup> segmentation assessment. Indeed, Chicco et al. (2020) indicates that the MCC, is a more reliable statistical rate [in the context of binary classification] which produces a high score only if the prediction obtained good results in all of the four confusion matrix categories (true positives, false negatives, true negatives, and false positives), proportionally both to the size of positive elements and the size of negative elements in the dataset. MCC indicates the correlation between the predictions and the reference data, and ranges from  $-1$  to  $1$ , with  $1$  being the perfect correlation. It is calculated as :

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP) * (FP + FN) * (TN + FP) * (FN + FN)}}$$

## 2.7 Related work

The scientific literature contains numerous studies on the extraction of agricultural fields, the delimitation of borders and the mapping of crops. It would be very long and inappropriate

<sup>1</sup>Field - non-field, boundary - non-boundary

to list all of the existing studies in the domain. Therefore, hereafter are listed some recent researches using state-of-the-art technologies in the field extraction domain.

Persello et al. (2019) is one of the first to use DL for agricultural field delineation from satellite images in the smallholder agriculture context. They used a SegNet model to detect field boundaries with panchromatic images from WorldView-2 and -3 satellites of Very High Resolution ( $0.5m$ ) in both Nigeria and Mali. The post-processing was realised by a Watershed Transform followed by a clustering technique to merge the most similar adjacent regions. Their results were promising, with F-scores higher than 0.7 and 0.6. Their proposed technique competes favourably against state-of-the-art computer vision contour detection algorithms in terms of accuracy. However, the ROIs were only respectively  $6km^2$  in Nigeria and  $30km^2$  in Mali, which cannot capture all the differences in land cover and topography, that are sometimes massive in those countries. Zhenrong et al. (2019) developed in the same year another DL semantic segmentation model based on DeepLabv3+. Their study only use the RGB bands from WorldView-2 with very high spatial resolution ( $1m$ ), on a  $1.500km^2$  ROI located in China. The agricultural system can be considered as smallholder agriculture, as the mean field area is 2 ha. Using some specific zones for validation purpose, they achieved Kappa = 0.9 and an OA of 95%. Their results outperformed every traditional machine learning methods, and other standard DL models (SegNet, UNet, PSPNet, DeelLabv2). Waldner et al. (2020) used their model developed in Diakogiannis et al. (2020) to realise a large field extraction in South-Africa. They used 10 Sentinel-2 tiles with  $10m$  resolution to train a ResUNet-a model, followed by a watershed transform for the post-processing. They achieved state-of-the-art results, with a MCC over 82%, and their model showed good space and time generalization capacities. As the field sizes were large with an average of 13 ha, they suggest it would be interesting to test the model in the smallholder agriculture context. Another type of interesting models are the Mask R-CNN (He et al., 2018), usually used to perform object detection, but have the potential to produce closed boundaries in a single step process. Zhong et al. (2023) introduced the MSMTransformer network, based on Mask2Former, to perform field extraction. Using Jilin-1 imagery with a spatial resolution of around  $1m$ , they achieved very promising results, proving that Transformer architecture has excellent performance in field extraction from RS images.

# Chapter 3: Objectives

## 3.1 General context

In the framework of collaboration between the Institut National de la Statistique du Mali (INSTAT, [2023](#)) and UCLouvain, the goal is to achieve a better understanding of the agricultural dynamics in this fast-growing region of the sub-Saharan region and improve the agricultural statistics. Achieving the cartography of the agricultural fields throughout the whole country would be a huge improvement to the food security. It would allow to be way more precise in the area estimation and yield predictions thanks to the superiority of vector layers on raster layers in biophysical variable retrieval and surface measurements respectively. A newly acquired set of reference agricultural field delimitation dataset (2021) of  $16.000km^2$  has been prepared using VHR visual interpretation by the Malian partner team of the Institut d'Economie Rurale (IER, Bamako) in order to make the first step in that direction.

## 3.2 Objectives

The overall aim of this research is to evaluate the performance of a field extraction methodology in the context of smallholder agriculture using a DL model on large-scale HRV satellite images. The first specific objective is to assess the field boundary accuracy obtained from a trained DL model from monthly Planet mosaics. The second specific objective investigates the model portability over a different region and the possible improvements by adjusting various components of the extracting field boundary models, including the quality of the input data and the imagery's spatial resolution.

The methodology will proceed with a calibration process, including the selection of the monthly mosaics and the tuning of hyperparameters, a validation process to qualify the model calibration, and post-processing optimization. The careful selection of segmentation performance metrics for assessing model results is also crucial, as explained in section [2.6](#). Due to the limitations of the reference dataset, a reasonably reachable target would be a dataset with a good prediction with respect to field location but not too severe regarding boundary accuracy.

# Chapter 4: Materials and methods

This chapter extensively describes the methods and input data selected in this research. The input data and its sources will be discussed, the pre-processing and the chosen methodology will be justified, and the additional experiments will be described.

DL terminology was employed in the redaction of the thesis. During the training phase, the deep learning model learns from the labeled data to optimize its parameters so that it can make accurate predictions on the training dataset. After each epoch, the trained model's performance is evaluated using the validation dataset. It is used for hyperparameter tuning and model selection. Finally, the testing set is a separate dataset used to evaluate the model's performance on unseen data, and simulate how well the model will perform in real-world scenarios.

## 4.1 Materials

In this section, the Regions Of Interest (ROI) will be first introduced (section 4.1.1) to give a brief description of the geographical context. The reference dataset will then be presented, along with its associated methodological problems due to the collection process (section 4.1.2). The satellite data that served as primary input and the joined pre-processing steps will be addressed next (section 4.1.3). The additional images used for the alternative experiments were obtained through different processes and will also be presented in the section 4.1.4.

### 4.1.1 Region of interest

Two ROIs are defined, based on the reference datasets provided by the Institut d'Economie Rurale (IER, Bamako). Both are located in the southern part of Mali, as shown on 4.1. The main ROI is located in the south of Mali, on the border between the Koulikoro and Sikasso regions, with approximately 95% of the ROI falling within the Sikasso region. It is a square of  $1.600 \text{ km}^2$  with the center located approximately  $50 \text{ km}$  west of the city of Koutiala, in the administrative circle sharing the same name. Koutiala is a major center for cotton production in Mali, mostly used for export purposes. Aside from that, various cereals are cultivated, such as millet, sorghum, and maize (Lambert, 2018). Some villages are located within the ROI, but none of them exceeds a few tens of thousands of inhabitants (Mali: Institut National de

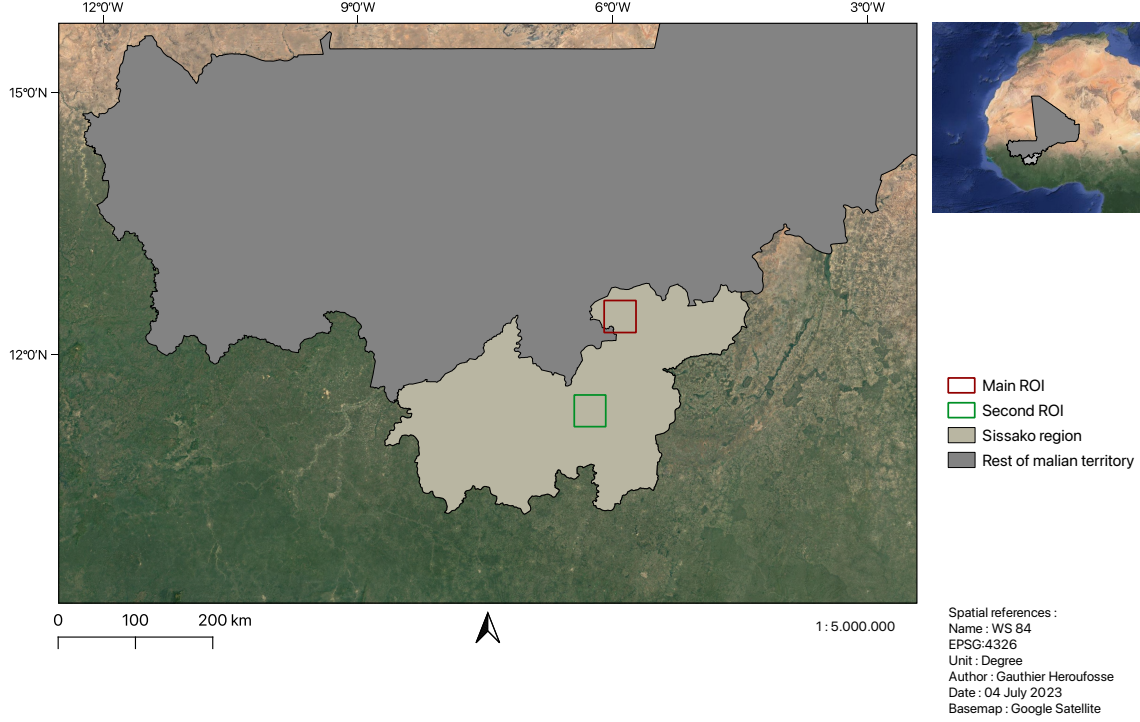


Figure 4.1: Location of the ROIs inside the malian territory.

la Statistique, 2012). The figure 4.2 shows the main ROI on the Malian territory.

Geographically, the area is situated in the Sudanian Savanna Zone, which serves as a transitional area between the Sahara Desert to the north and the more forested regions to the south. The region experiences a dry season from November to April, followed by a rainy season from May to November, which significantly impacts agricultural practices in the area. According to the Köppen-Geiger climate classification, the ROIs fall into the BSh classification. Figure 4.3 illustrates the average precipitation distribution over the year in the city of Koutiala. The onset of the cropping season is marked by the first rainfall, which can vary with the years. As only the 2021 year is analyzed here, it is important to take the local variation that occurred specifically that year into account. The USDA (2023) data show that the 2021 season started on the 20th of March and ended around the 20th of November (see Figures A.1 and A.2 in Additional material and methods figures). It is important to note that the landscape is a mixture of flat plains, gentle slopes, and occasional small hills, which play a role in shaping the agricultural context of the region (Lambert, 2018). The ROI is crossed by numerous tributaries of the Niger River, such as the Banifing-Lotio River. These small rivers have a major influence on agriculture, with farmers flooding the region to cope with the lack of rainfall.

The second ROI is used for validation purposes in an additional experiment. Located 130 km south of the main ROI, which is also 60km west of the city of Sikasso, the field density



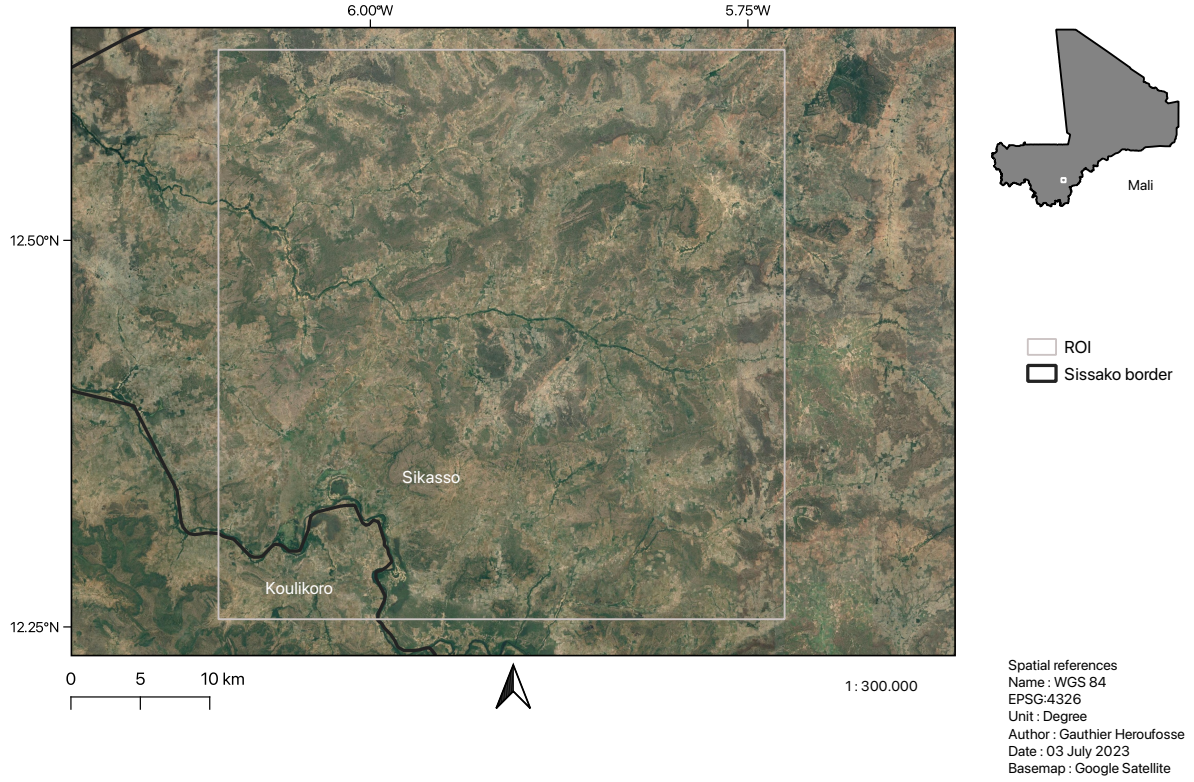


Figure 4.2: Location of the main ROI. The border between the Sikasso and the Koulikoro region is visible in the bottom left corner inside the ROI.

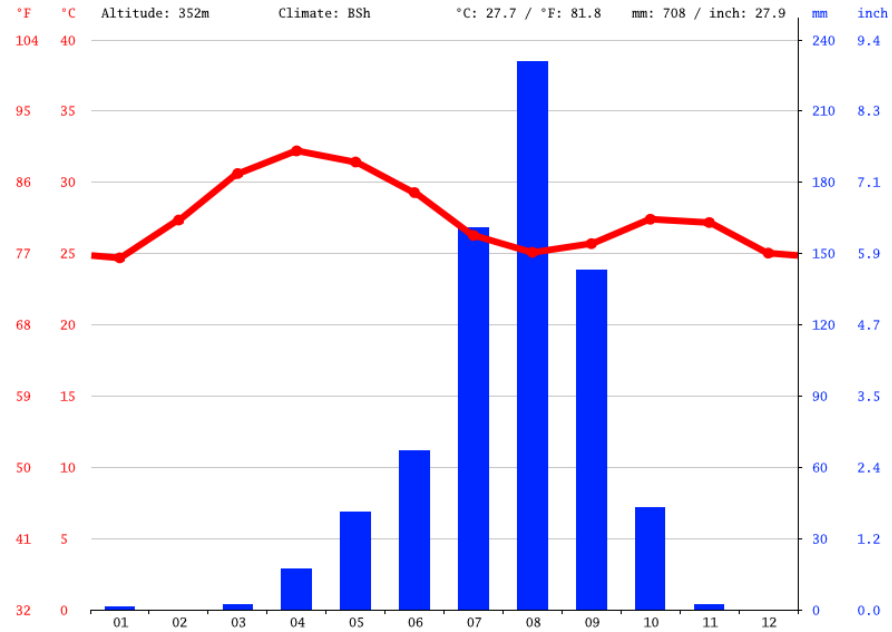


Figure 4.3: Average monthly precipitation and temperature in Koutiala, from *Climate data for cities worldwide - Climate-Data.org* (2023).

is lower, and a few larger cities are present in the area. This area will be used to assess the portability of the model within a nearby area, between two relatively similar landscapes. It will be detailed later in Section 4.2.4.3.

### 4.1.2 Reference datasets collection and quality control

The reference dataset used for this research consists of fields manually digitized in the ROI presented before, obtained from the Institut National de la Statistique of Mali (INSTAT, 2023). The 50.230 fields located in the  $1600 \text{ km}^2$  square were delineated using VHR images, mostly from Google Maps as a primary source and from Bing as a second source. The images from Google Maps in the ROI are from 2021, but the exact satellite source is not made public. These are mosaic images that were recomposed using images from different periods of the year, to achieve the best visual output. However, it is actually not possible to find the exact date on the Bing mosaics. After a quick visual analysis, it seems obvious that Bing images are older than Google images. As a significant part of the reference fields were also delimited using Bing images as a primary source, it results in some spatial inconsistencies in the delimitation due to the year difference. For the sake of efficiency, the team in charge of the field delineation process proceeded by square of  $5 \text{ km}$  side, with a few different interpreters in charge of all the squares. The intermediate  $25 \text{ km}^2$  squares were then not merged properly, resulting in straight ends of all the field boundaries every  $5 \text{ km}$ , both horizontally and vertically. Furthermore, it is quite challenging to harmonize the visual interpretation of different interpreters with different perceptions of what is a field boundary or not. An unambiguous delineation would have required buying a VHR imagery time series to capture the seasonal contrasts. In some regions, all differences are considered, and the fields are then very small, where other interpreters preferred to keep larger fields. The figure 4.4 illustrates those differences by way of example.

These problems are unfortunately impossible to completely solve, as it would require analyzing the  $1.600 \text{ km}^2$  ROI again. To partially address the issue, two processes were implemented. First, all the intermediate perimeters of  $25 \text{ km}^2$  were manually merged together. The ROI was extensively scanned every  $5 \text{ km}$  both horizontally and vertically on the  $40 \text{ km}$  length, and every unnatural field boundary found was edited to match the Google satellite image correctly. Some areas in particularly bad shape were also manually corrected, such as a large region where no field was referenced at all. After the application of those modifications, a new total of 52.107 polygons was obtained. The fields have a mean of  $12.467 \text{ m}^2$ , i.e. 1,2 ha, with a standard deviation of  $12.410 \text{ m}^2$ . This is mostly due to a few really big fields in the ROI, as the interquartile range is  $11.154 \text{ m}^2$ . It is important to understand that the 'corrected' reference dataset is still not complete. This impacts the generalization capacity of the model outside its training area. Indeed, in the context of DL, every mistake in the reference weights doubles, as the model not only misses the correct label but also learns that it is actually part of another class. This leads to the overfitting of the model to corrupted labels, due to its capability of learning any complex function (Song et al., 2022). The complete quality-controlled reference dataset is presented on Figure A.3 in Additional material and methods figures, while the figure 4.5 is a zoom for visualization purposes.

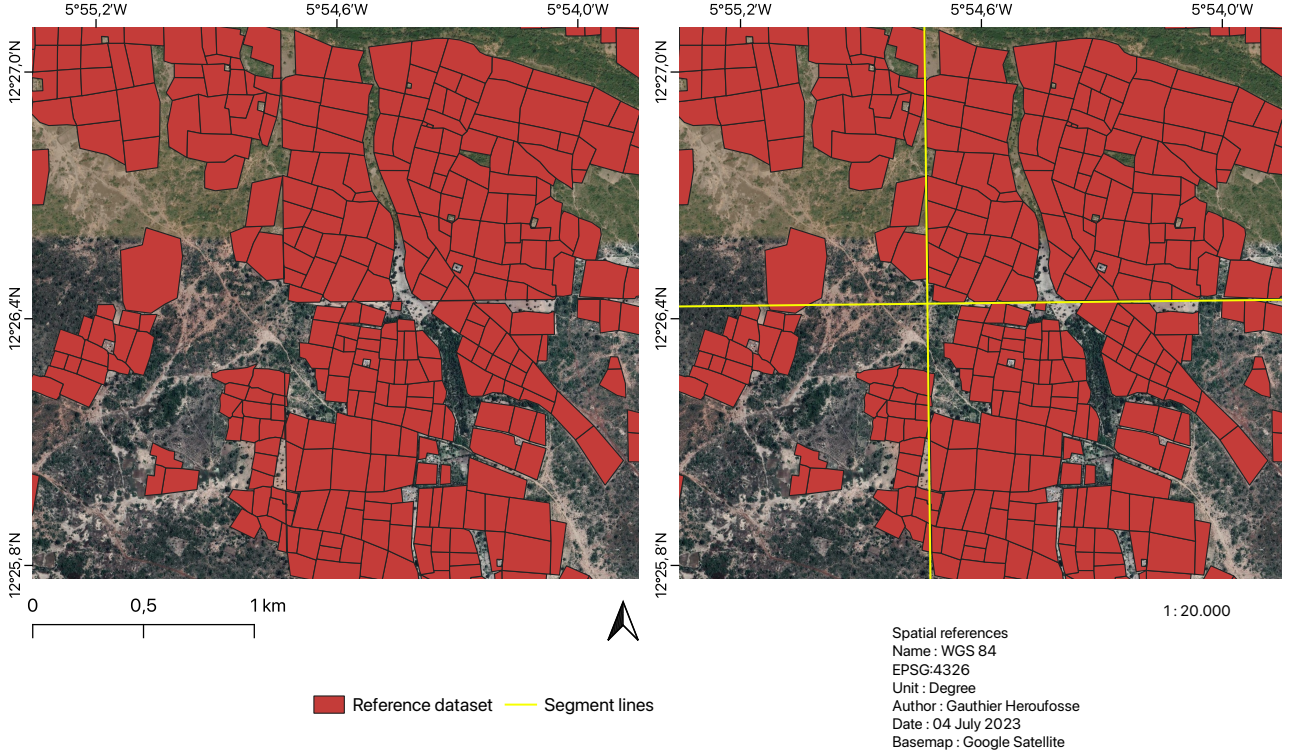


Figure 4.4: Illustration of the problem related to the split between sub-perimeters for the data segmentation. The yellow segment lines trace the delimitation between those imaginary squares.

Secondly, a smaller area was more precisely analyzed, and each field was examined to ensure that the boundaries were correctly drawn. This small area of interest was used for the calibration analysis and is referred to as the Quality Set. The set comprises 1.903 fields covering an area of  $50\text{km}^2$  for training and 819 fields covering an area of  $19\text{km}^2$  for testing purposes. Its location has been chosen to represent the reference data as faithfully as possible, with the presence of forests, villages, and water bodies. As for the reference dataset presented before, the complete Quality Set is presented on figure [A.4](#) in [Additional material and methods figures](#) while the figure [4.6](#) is a zoom for visualization purposes.

The quality-controlled reference dataset was then split into a training, a validation, and a testing set. 12.563 fields were selected to compose the testing set, which constitutes 24% of the dataset, and the remaining features correspond to the training and validation sets (training data). The reference dataset was rasterized at the same resolution as the input Planet satellite data (see Section [4.1.3](#)) to match the pixel size. It results in  $>75$  million reference pixels. This so-called 'reference mask layer' is then used during the training process to create three other reference raster layers: a binary layer of the extent of fields, a binary layer of the field bound-



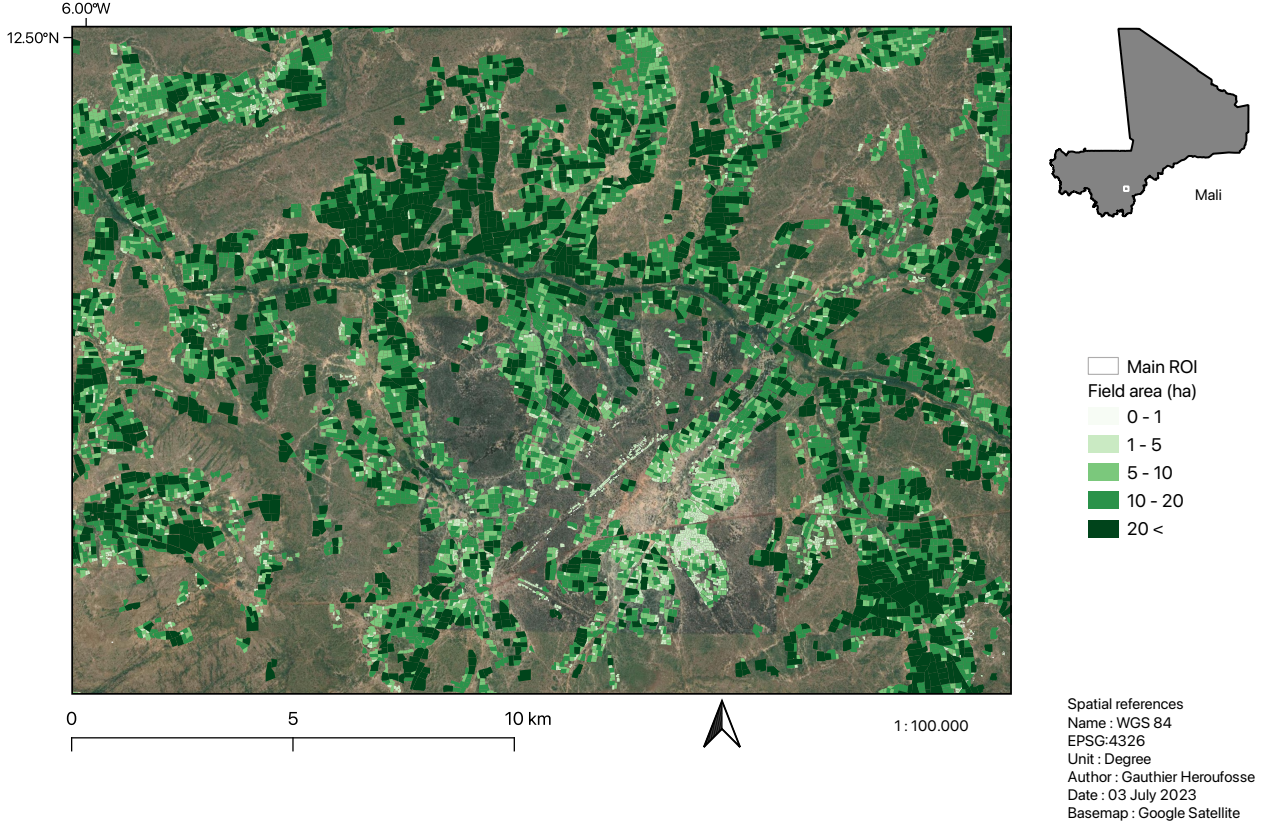


Figure 4.5: Zoom on the quality controlled reference dataset displaying the field size of each individual polygon. The complete dataset is available in [Additional material and methods figures](#).

aries, with a 10-m buffer applied, and a distance transform layer with respect to the closest boundary layer. Distances were normalized per field so that the largest distance was always one within each field. Those three layers are used to calculate the loss during the model training.

A second reference dataset was collected to cover the second ROI. This reference dataset will only be used for validation purposes in one of the additional experiments. It is made up of 31.527 fields manually digitized, as for the first reference dataset. It suffers basically from the same issues as the first reference dataset, but no correction was applied as it would request more days of work than were actually available.

### 4.1.3 Planet Data Pre-processing

The VHR images were collected from the Planet Basemaps program (Planet, 2023). It consists of small mosaic images of the year 2021. Planet images were chosen thanks to their free availability and attractive spatial resolution (4.77m). The images are composed of 5 bands (Blue, Green, Red, NIR, and an alpha band for transparency) with surface reflectance values. As they are not daily images but mosaics, the images are mostly cloud-free. The assembling

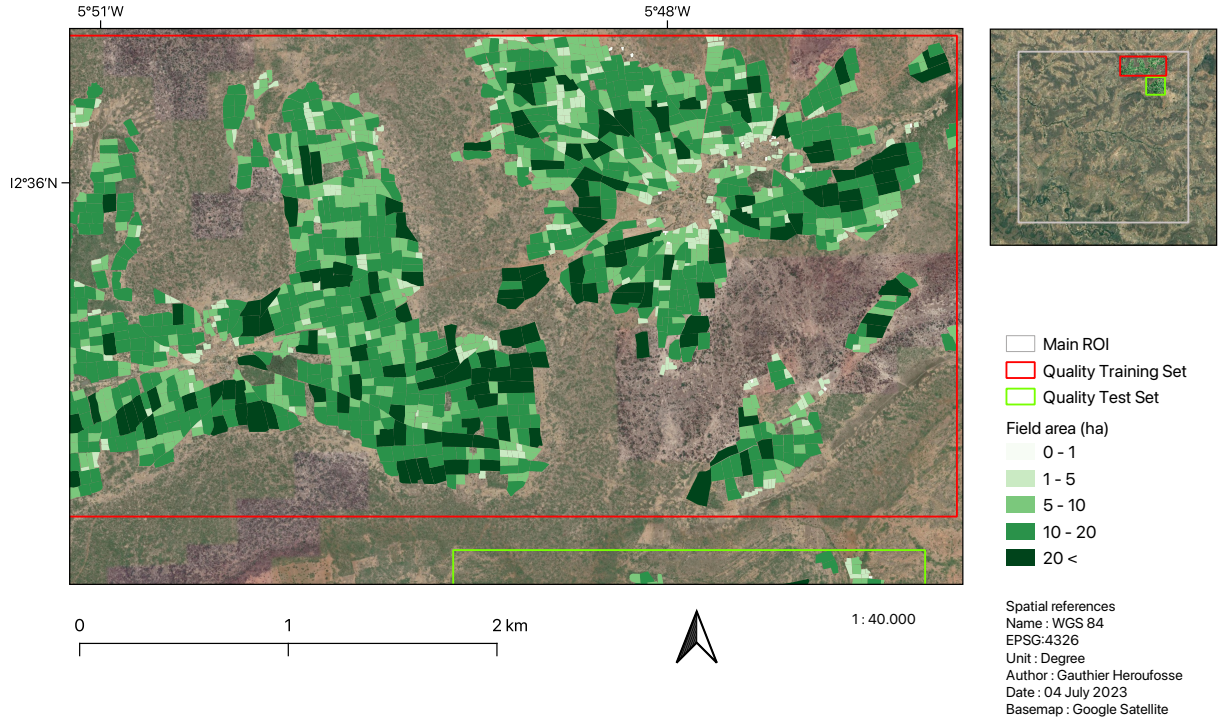


Figure 4.6: Zoom on the Quality set. The complete Quality set is available in [Additional material and methods figures](#).

and correcting process of the mosaics is not detailed by Planet, and some cloud aggregates are still visible on some images. An example of those aggregates can be seen in figure [4.7](#).

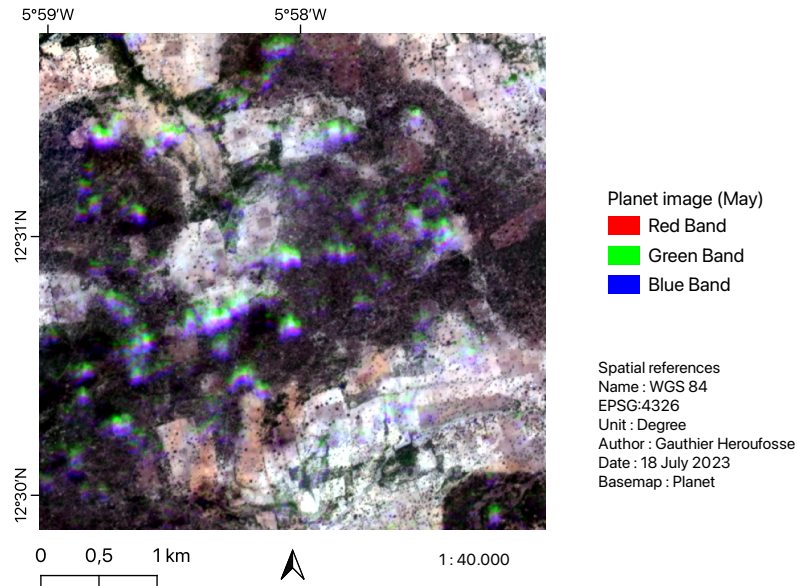


Figure 4.7: Illustration of the remaining aggregates present on the Planet satellite images.

The first preprocessing steps consisted of merging the small images together into a single image for each month, covering completely the ROI, and then cropping it so that it matches exactly

the extent. Then, the extents of the training and testing sets were used to crop the full image to have one training and one testing image per month. Once the alpha band has been removed, the images will be concatenated to form stacks of three-month images (both training and testing stacks). As it is not possible to process all the images at once due to the GPU's limited memory, the training stack was divided into a set of  $256 \times 256$  pixel images, each with the right amount of bands (12 bands for a three-month stack). Input images on the borders were discarded for the training process due to their incompleteness. The total number of images for the baseline method (see Section 4.2.2) was 12.672 images, split with 10% being the validation images and the rest being the training images. As a reminder, training images are used to train the deep neural network, and validation images are used for parameter tuning within the training process. The last preprocessing step is the element-wise standardization of the input data. The standardization of the data prior to the training of the neural network is necessary, because neural networks are sensitive to the scale of the input variables. Wan (2019) found that gradient optimization methods converge faster when features have a zero mean and unit variance.

Data augmentation was also used to add some extra variety to the input data. Two operations are conducted at random: each pair of image and ground truth is zoomed in or out, and rotated with a random center. A reflector padding is used to fill in the parts of the image that are left out of the frame after the transformation. The data augmentation methodology comes from Diakogiannis et al. (2020). The figure 4.8 illustrates the data augmentation result.

#### 4.1.4 Additional images

Some extra images needed to be collected or processed to complete the additional analysis (see Section 4.2.4). The first analysis assesses the possibility of using VHR images for field extraction in smallholder agriculture. Maxar images (Maxar, 2023) from 2021 with 1m spatial resolution were used. Despite the advantageous resolution, some drawbacks could affect the result of the analysis. The absence of the NIR band is seen as a major problem, as this band forms the basis of several spectral indices that can be used to distinguish vegetation from other types of land cover. The image is not a single image, but rather a mosaic made of multiple small images with different dates. This is done to obtain a cloud-free image, but at the cost of abrupt changes between two neighboring zones. The training zone is the same as for the Quality Training Set. However, the testing zone is different from the Quality Test Set, as the mosaic in that location was not usable. The last analysis, exploring the portability of the model in nearby areas, needs the same Planet image type as the one used for training. Basemap Planet images covering the second ROI were collected.



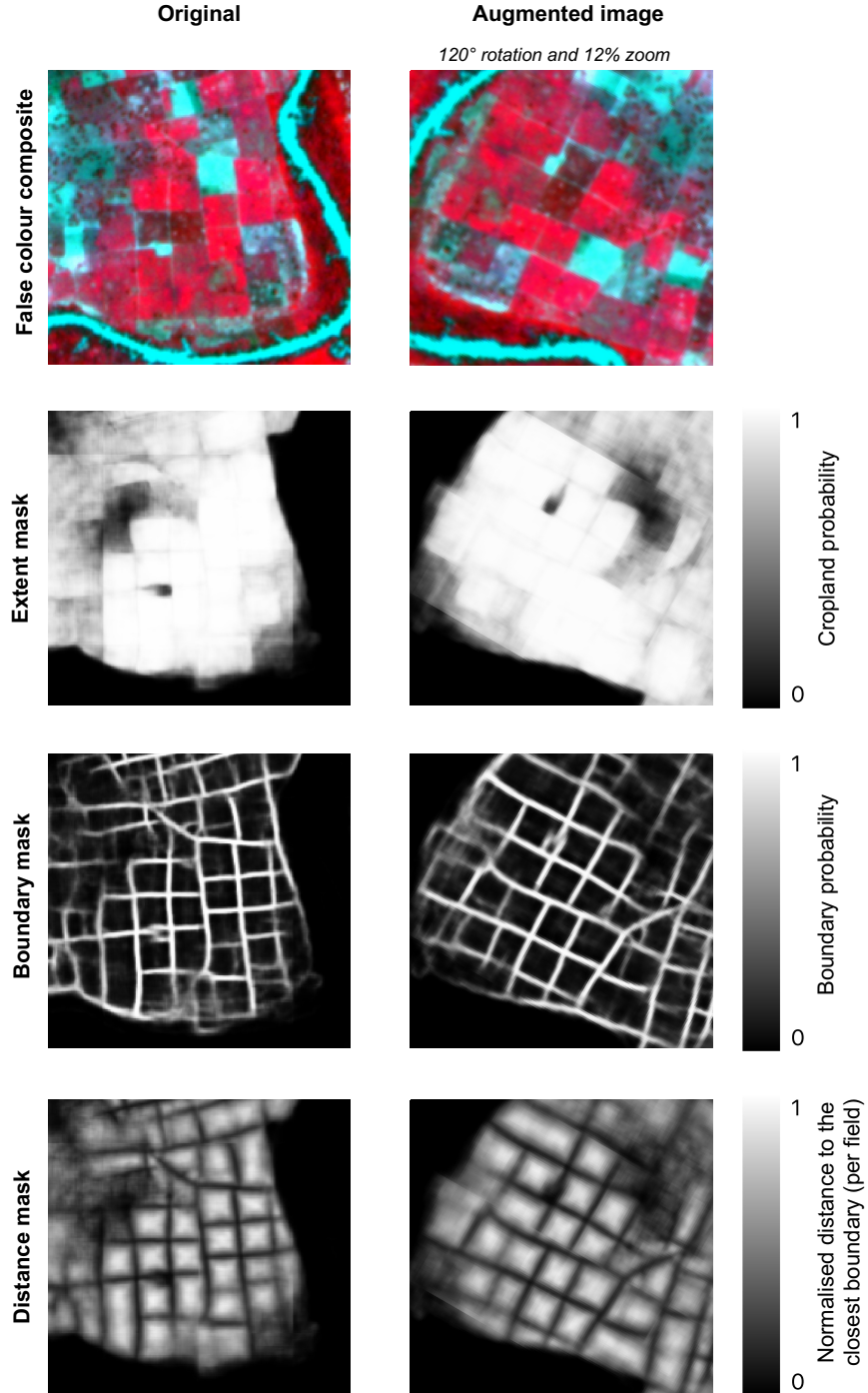


Figure 4.8: Data augmentation example. The left image represent the original, and the right the corresponding rotated and zoomed in image.

## 4.2 Methods

The section discusses the selected model and the methodology followed to tackle the objectives. An in-depth explanation of the model and its architecture will be given first (section 4.2.1). Several experiments were conducted to evaluate the pertinence of the model to extract field boundaries in a smallholding agricultural landscape. The baseline method (section 4.2.2) uses an association of three Planet mosaics and explores the influence of the monthly associa-

tions chosen, as well as the variation of the hyperparameters chosen, according to a set of pixel- and object-wise metrics, in order to maximize the concordance between the reference and the predicted fields. Additional experiments were conducted in order to generalize the approach to various data types, and assess the portability of a field extraction DL model to relatively nearby locations in Mali. They are presented in section 4.2.4. The list of software and packages used for the fulfillment of the thesis is available in Section 4.2.5. The section ends with a flowchart resuming the complete methodology, from the data preprocessing to the accuracy assessment.

### 4.2.1 Model selection

The model chosen is the ResUNet-a model developed by Diakogiannis et al. (2020). Indeed, as it has obtained state-of-the-art results in a large-scale application in the field extraction process, the evaluation of its generalization capabilities to smallholder agriculture is an area of interest. As briefly introduced in the section 2.7, ResUNet-a is a FCN. The architecture is complex and made of various blocks designed to achieve a specific operation. The model is made out of the following components:

1. UNet backbone architecture (section 2.4.2), i.e., the encoder-decoder paradigm ;
2. Residual blocks as building blocks of the UNet architecture (He et al., 2016) ;
3. To enhance comprehension at different scales, each residual building block incorporates multiple parallel atrous convolutions (Chen et al., 2017) with varying dilation rates ;
4. Pyramid Scene Parsing pooling (PSPPooling) Zhao et al., 2017 to include contextual information ;
5. Multitask learning, the algorithm simultaneously learns four complementary tasks.

Diakogiannis et al. (2020) introduced two ResUNet-a models, the ResUNet-a D6 and the ResUNet-a D7. The difference lies in the number of residual building blocks: ResUNet-a D6 has six residual blocks of them followed by a PSPPooling layer, where ResUNet-a D7 has instead seven residual blocks followed by either a Max Pooling layer (D7v1) or a smaller PSPPooling layer (D7v2). This comes at the cost of a larger number of parameters (160 million for the D7 vs. 52 million for the D6). Waldner et al. (2020) demonstrated that the results they obtained in their field extraction process, which is a close procedure from what is implemented here, were slightly better with the ResUNet-a D7(v2) model. Therefore, only this model was retained for further analysis.

The layer types in the ResUNet-a model have the following function, as described by Diakogiannis et al. (2020) :

The **Input layer** takes the input image, which is a  $256 \times 256$  raster with a mutable number of bands.

The **Conv2DN layer** consists of a 2D convolution layer with a kernel size of 3 and padding of 1, followed by a batch normalization layer. Batch normalization is an effective technique used to address the internal covariate shift problem in neural networks. It works by normalizing the output of the preceding activation layer, subtracting the batch mean, and dividing it by the batch standard deviation. This process significantly improves the speed, performance, and stability of artificial neural networks (Szegedy et al., 2015).

The **ResBlock-a layer** implements the principles of residual units (He et al., 2018), which involve incorporating skip connections between layers to enhance the information flow from the initial to the final layers. Rather than employing a single residual branch composed of two consecutive convolution layers, ResBlock-a employs up to four parallel branches with progressively larger dilation rates. This enables simultaneous processing of the input in multiple fields of view. The selection of dilation rates is determined by the size of the input feature map.

The **PSPPooling layer**, inspired by the pyramid scene parsing pooling technique (Zhao et al., 2017), focuses on capturing contextual information. It achieves this by employing maximum pooling at four distinct scales. The first scale involves global max pooling, while in the subsequent scales, the feature map is divided into four equal areas, and max pooling is performed independently in each of these areas. This process continues for the remaining two scales. Consequently, this layer effectively encodes information about prominent contextual features across various scales.

The **Upsample layer** involves two main steps. Firstly, it performs an initial upsampling of the feature map using interpolation. Subsequently, a Conv2DN layer is applied. During the upsampling process, the size of the feature map is doubled, while the number of filters in the feature map is halved.

The **Combine layer** accepts two inputs, each having feature maps with an equal number of filters. It combines these inputs by concatenating them and generates an output that preserves the same scale and number of filters as each of the input feature maps.

The **Output layer** performs multiple tasks with conditioned inference and generates four output layers: the extent mask, the boundary mask, the distance mask, and the reconstruction of the original image in the Hue-Saturation-Value (HSV) space. The network graph is designed to leverage the inference results from previous layers. The process begins by combining the first and last layers of the UNet. The distance mask is the initial output, computed without utilizing

PSPPooling. Next, the boundary mask is produced by combining the distance transform with the last UNet layer using PSPPooling. The extent mask is generated by combining the distance mask and boundary mask with the output features. Additionally, a parallel fourth layer reconstructs the initial input image, which helps reduce model variance (Diakogiannis et al., 2020).

## 4.2.2 Baseline method

The baseline method aims to achieve the best possible prediction result on the Testing Set, according to both a visual appreciation and a set of chosen metrics (presented in the section 4.2.3). The common setup includes a ResUNet-a D7(v2) model initialized with the Xavier initializer (Glorot et al., 2010), and the Adam optimizer (Kingma et al., 2017). Models were trained using Tesla A100 GPUs for a maximum of 80 epochs, depending on the analysis. The batch size was set to 16 due to memory limitations. The Tanimoto loss function used for the training is presented in detail in the next section. Once the calibration analysis was done, the last model was trained on the complete Training set with the best training setup. See Hecht-nielsen (1992) and Chauvin (1995) for explanations of the process of training a CNN.

### 4.2.2.1 Calibration analysis

The process asks for an extensive calibration analysis, aiming to choose the best input images and tune the hyperparameters. To do so, the Quality Training Set was used, as the reference dataset is completely reliable, and it is approximately 50 times faster to train a model on this set than on the full Training Set. The selection was made on the basis of the validation metrics, namely the Extent Accuracy, the MCC Extent, the Boundary Accuracy, and the MCC Boundary. MCC and Accuracy are chosen because they are, respectively, the most robust and most used metrics in binary segmentation assessment. The calibration analysis was carried out in the following order:

1. Find the best monthly mosaic association ;
2. Identify the best Learning Rate (LR) approach ;
3. Set the best loss combination.

The parameter settings that gave the best result according to the validation metrics were then retained for subsequent evaluations.

## 1. Selection of monthly mosaic set

This first analysis looks to select input data that gives the best results. A choice needed to be made, as the use of more than 12 bands was not possible as the memory needed jumped above 50 *Gb* of storage, therefore making the training impossible to complete with our available resources. As a reminder, every mosaic is composed of 4 bands (RGB and NIR), which means that up to 3 months could be associated. All twelve monthly mosaic images were available, except for April, due to the many remaining clouds on the mosaic. After a visual examination, the analysis was carried out on the months between May and November, as a combination of those months seems the most appropriate to discriminate crops for the rest of the land cover. It is directly related to the seasonality presented in section 4.1.1. The use of multiple months was found to be relevant, as more information is available with 3 images (12 bands) than on a single image (4 bands). The validation metrics gave up to a 0.07 difference between MCC Extent and a 0.04 difference between MCC Boundary. However, this addition comes with a new issue: the reconstruction of the HSV image uses only three bands. Although the HSV image is not mandatory to achieve the segmentation process, Diakogiannis et al. (2020) showed that it helps to reduce the model variance and make sure that all fine details are preserved. An assessment was thus made to see if the result without the HSV layer output was indeed helpful in our study case, and we concluded it is useful (see section 5.1 for the results and explanations about the loss further down here). A selection was arbitrarily set to use the 3 NIR bands (4th, 8th, and 12th) for the image reconstruction, assuming that those bands carry the most useful information to perform an accurate segmentation. The models were trained for 20 epochs with a constant LR of  $10^{-3}$ .

## 2. Learning Rate tuning

The second analysis covers the influence of the choice of the initial LR on the model results. Bengio (2012) indicates that this is often the single most important hyperparameter, and one should always make sure that it has been tuned. The analysis is divided into three parts: the measure of the variation between two runs with the exact same model, the selection of the appropriate starting learning rate, and the relevance of the addition of a Weight Decay (WD) approach. The first step is to measure how much the results can vary between two runs of the same model. Indeed, optimization methods in DL often involve stochastic algorithms, subject to random variation. The Adam optimizer makes no exception to the rule, and variability arises from the combination of mini-batch sampling and adaptive learning rates for individual model parameters (Kingma et al., 2017). Plus, the Xavier initialization induces a random initialization of the weights of the neural network connections, with values drawn from a specific distribution.

The second step consisted of training multiple models with various starting LR in order to



select the one to keep for further analysis. Three different LR's were initially tried:  $10^{-3}$ ,  $10^{-4}$  and  $10^{-5}$ . A consecutive study around the best LR ( $10^{-3}$ ) was then conducted to take the analysis a step further. Five learning rates between  $5 * 10^{-3}$  and  $5 * 10^{-4}$  were used to make the final starting learning rate choice.

The last step consists of analyzing the relevance of using a Weight Decay (WD) approach in the training process. The method adds a weight decay parameter to the learning rate in order to decrease the neural network weight and bias values by a small amount in each training iteration when they are updated. Four different settings were tested with the same factor (0,3162)<sup>1</sup>: after 10 epochs, after 20 epochs, after 40 epochs, and no decrease at all.

### 3. Loss Analysis

This last analysis explores the influence of the Loss composition on the results. Diakogiannis et al. (2020) introduced the Tanimoto loss as a loss function based on the DICE coefficient. They showed that, thanks to gradient flow lines that are geodesics, the Tanimoto loss achieves the best results among the loss' family based on the DICE coefficient, in the context of DL models training. As for Waldner et al. (2020), our loss function relied on the Tanimoto loss, defined as the average of the Tanimoto loss  $T(p, l)$  and its complement  $T(1 - p, 1 - l)$ , such as :

$$\tilde{T}(1 - p, 1 - l) = \frac{T(p, l) + T(1 - p, 1 - l)}{2}$$

with

$$T(p, l) = \frac{\sum_i p_i l_i}{\sum_i (p_i^2 l_i^2) - \sum_i (p_i l_i)}$$

where  $\mathbf{p} \equiv p_i \in [0, 1]$ , represents the vector of probabilities for the  $i$ th pixel, and  $i$ th are the corresponding ground truth labels. For multiple segmentation tasks (here Extent mask, Boundary mask, distance to boundary layer, and HSV reconstitution layer), the complete loss function is defined as the weighted average of the loss of all tasks :

$$\tilde{T}_{MST}(p, l) = \frac{w1 * \tilde{T}_{Extent} + w2 * \tilde{T}_{Boundary} + w3 * \tilde{T}_{Distance} + w4 * \tilde{T}_{HSV}}{\sum_i w_i}$$

with weights initially set to one.

As previously discussed, the use of multiple images in a single stack for training purposes raises questions about the HSV reconstruction. Hence, a comparison between a model trained with and without the use of HSV was made to assess the importance of HSV in our context. As previously discussed, the HSV used in our case is made up of the three NIR bands (from

<sup>1</sup>The factor is such that every two decreases, the LR drops by 10%

each of the monthly mosaics). The resulting loss function without the HSV layer was defined as:

$$\tilde{T}_{MST}(p, l) = \frac{1 * \tilde{T}_{Extent} + 1 * \tilde{T}_{Boundary} + 1 * \tilde{T}_{Distance} + 0 * \tilde{T}_{HSV}}{3}$$

Secondly, a double weighting of the Extent and the Boundary was also tested, to switch the focus on either of those outputs. Indeed, despite the boundary having the largest weight in the loss calculation, the MCC values were still half as high as the Extent MCC. Two different manipulations were tried to correct those differences: doubling the weight of the Boundary to increase the training focus on this task, and doubling the extent focus to balance the distribution of loss. The resulting loss follows the same scheme as the loss without HSV, with varying weight coefficients.

#### 4.2.2.2 Inference methodology

Once the model is fully trained, the prediction on the testing stack can be realized. The model can be applied to any 256x256 image with a simple forward pass through the network. As the testing stack is not divided into 256x256 images, a sliding 256x256 window is applied in practice. However, convolutional neural networks use contextual information for prediction, which means that the classification accuracy of a pixel will depend on its location. This results in homogeneous 256x256 windows on the predicted image, without any relation between neighboring windows. To compensate for this location-related issue, a 64-pixel slide in every direction was applied. The final class probability vector is obtained using the average of all prediction views. The same principle is implemented on the side of the image, with a reflect padding at a size of 128 pixels (Diakogiannis et al., 2020).

#### 4.2.2.3 Post-processing

Although ResUNet-a learns to detect field boundaries, it is not capable of generating automatically closed boundaries, which is required to extract individual fields. The *watershed* method raises field boundaries and therefore lets us achieve instance segmentation and obtain individual fields. The method delineates individual fields by considering the outputs of ResUNet-a as a topographic map. Water is then filled into the map, starting at each local minima, until it reaches a certain height threshold. Fields stop growing in a direction when the threshold is reached or if a border hits the boundary of another object. The local minima are found using the distance map, the topographic surface is defined by thresholding the extent mask, and the boundary threshold is set by the boundary mask. The threshold selection was made using visual appreciation of the result, followed by a manual calibration around the initial

thresholds. The selection criteria for the process were the performance metrics that are presented in the next section (Section 4.2.3). To compute the metrics, fields smaller than  $1.000m^2$  were removed from both the reference and the prediction datasets, as well as fields larger than  $800.000m^2$ . It represents less than 1% of the polygons for the quality controlled reference dataset but around 15% for the prediction dataset. In addition, a 2-pixel buffer (around  $10m$ ) was applied to the reference boundary mask layer (see section 4.1.2) to reduce the strictness of the evaluation.

### 4.2.3 Performance evaluation

As explained in Section 2.6.2, the evaluation of a segmentation result is all but a trivial task. The optimal segmentation concept is subject to a very application-dependent appreciation, and should be defined for each application considering the needs and requirements of the research projects. In our case, as the reference dataset is not perfect, an optimal result would consist of an agricultural parceling that is not too severe about the under- and over-segmentation. Indeed, Section 4.1.2 introduces the several problems linked to the reference dataset and its lack of precise boundaries. Therefore, a realistic segmentation result would have a good prediction with respect to extent, but would not be too severe with respect to boundary accuracy. The evaluation will be both qualitative and quantitative. This would also allow to visually control the results.

The qualitative assessment of the results will consist of a close visual analysis of the tested area, under the prism of the following criteria:

- Completeness and False Positive ;
- Consistency and Coherence of the area ;
- Likelihood of the field boundaries ;
- Over and Undersegmentation.

Qualitative evaluation gives us a vision of the results that is partly distinct from the reference dataset, which in our case is not totally reliable. In the quantitative assessment, the evaluation of the quality of the extent and the boundaries has been separated, in order to have a distinct evaluation between the detection of the boundaries and the detection of the extent of the field. This gives a concrete view of the results concerning boundaries, which the model finds harder to learn. Model accuracy was evaluated using both pixel-based and object-based metrics. Those metrics were chosen to provide the most comprehensive information on the results and because of their high visibility for comparison with further studies. Pixel-based accuracy was computed

from the confusion matrix with the help of a few different metrics : MCC, the F1-score, Precision and Recall, and the OA (see Section 2.6.3).

Object-based accuracy was evaluated using a few different metrics. The first two metrics are, respectively, *Precision* and *Recall*, introduced earlier (Section 2.6.2). *Precision* is a measure of the under-segmentation, while *Recall* expresses the over-segmentation (Costa et al., 2018). Those two metrics are then combined in the  $F\text{-measure}_{PR}$  as follows:

$$F\text{-measure}_{PR} = \frac{1}{\alpha \frac{1}{Precision} + (1 - \alpha) \frac{1}{Recall}}$$

The third object-based metric we use is the  $RecRate_{soft}$ . Introduced in Pavlova et al. (2021), it ignores the errors due to over- and under-segmentation. It goes along with the FPR, which, as its name suggests, counts the number of fields from the prediction dataset that do not share area with the reference dataset. Both metrics have already been extensively described in Section 2.6.2.

## 4.2.4 Additional experiments

The previous research set the baseline method for field boundary extraction in the context of smallholding agriculture. This next section gathers some extra experiments trying different types of input data, as well as exploring the generalization capabilities of the baseline method in nearby space of the same variety.

### 4.2.4.1 Impact of the quality of the reference dataset on the segmentation results

This additional analysis examines the influence of the quality of the reference dataset on the segmentation results. As explained in the Section 4.1.2, the complete reference is far from being a reference on which we can 100% rely. To estimate how far this reference dataset deviates from the truth, the Quality Set was compared to the original reference segmentation. As for the baseline method, the performance evaluation will be both quantitative and qualitative. In addition, a prediction on the validation set will be made using a model trained on the Quality Set. This illustrates the variability of landscape architecture and therefore the importance of a large amount of training data.

### 4.2.4.2 Very High Resolution images in the context of smallholding agriculture segmentation

VHR images allow access to way more information about the image than HR images. One can easily distinguish the trees, the roads, and even the details on the buildings. A ResUNet-a

will be trained using Maxar images resampled to  $1m$  of spatial resolution on the Quality Set.

#### 4.2.4.3 Assessing the model portability within nearby area

Finally, we assessed the model’s generalization capabilities using the secondary ROI as a testing area. Planet images of the same resolution as for the baseline method were used to extract field boundaries with the best model resulting from the baseline method.

#### 4.2.5 Softwares used

The ResUNet-a model was implemented by Diakogiannis et al. (2020) using the *MXNET* deep learning library, under the GLUON API. The model was trained using one Tesla A100 GPU in the CISM facility (see Acknowledgements). Additional Python libraries were used for georeferencing manipulations such as *gdal*, *geopandas*, *osgeo*, *rasterio* and *Orfeo toolbox*. *Matplotlib* and *OpenCV* were used for visualisation and image processing. Finally, *QGIS* was used for visual analyses of satellite images.

### 4.2.6 Flowchart

The figure 4.9 shows a schematic overview of the whole methodology for the Basline method. The calibration analysis indicates multiple runs of slightly different models in order to achieve the best result. The process using the Quality Set is essentially the same, with the addition of an extra manual field boundary correction.

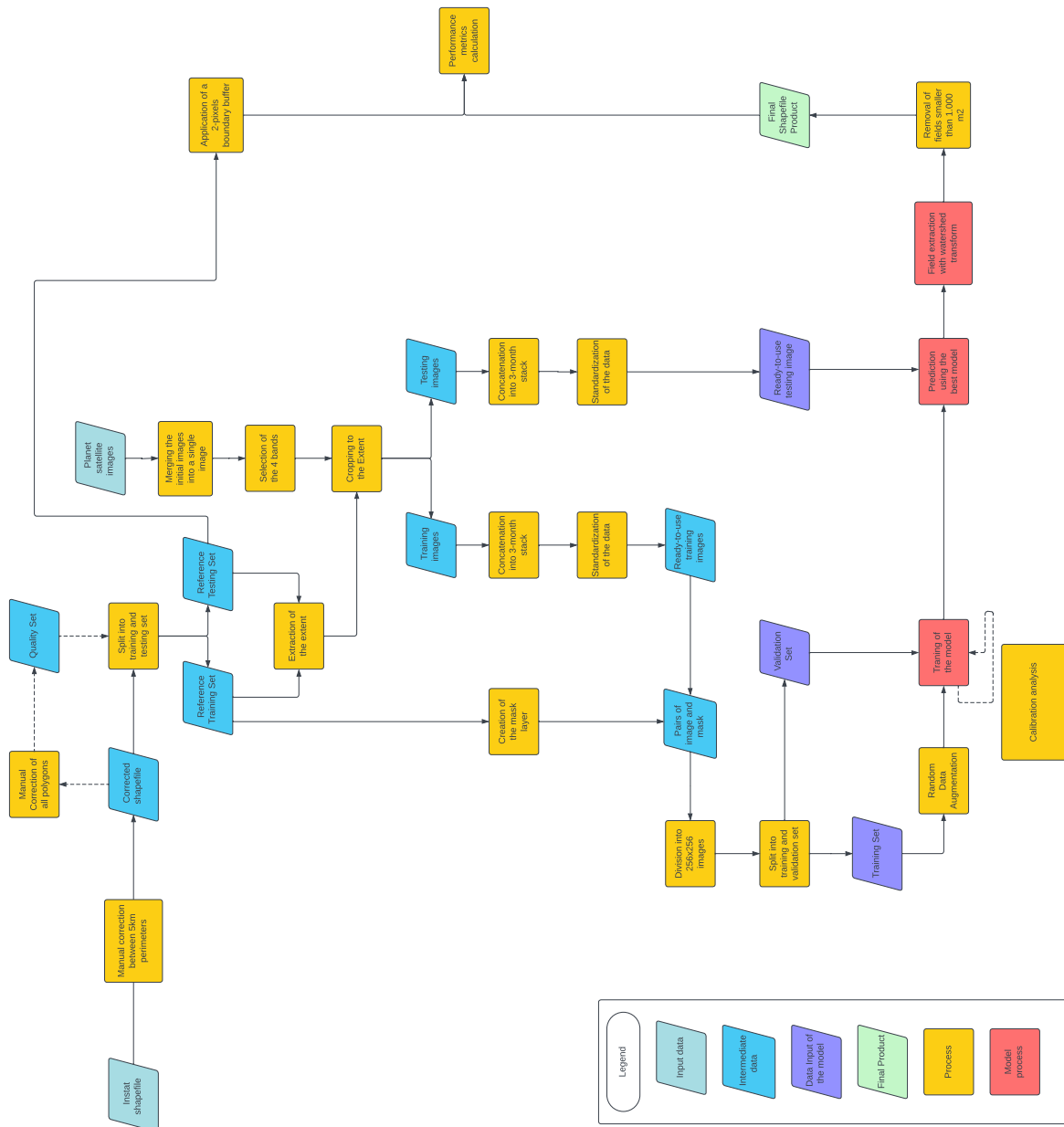


Figure 4.9: Flowchart of the complete methodology applied throughout this master's thesis.

# Chapter 5: Results and discussion

This chapter will gather all the results obtained with the methodology described in the previous section. The baseline method and its calibration will be addressed first, followed by the final results obtained with the best model implementation. The impact of the reference dataset will then be presented in the Quality Set assessment. Finally, the additional analysis shows the influence of changes in the data input and the generalization capabilities. The chapter will be concluded by a broad discussion of the results and some key elements to explore in further work.

## 5.1 Baseline method results

This section starts by giving the results of the calibration analysis to obtain the last implementation parameters. It is followed by the final product obtained from the baseline method and its accuracy assessment.

### 5.1.1 Calibration analysis results

The results from the calibration analysis will be detailed in the following section. Some graphs are present directly in the text, while the rest are displayed in Section [B](#): Appendix B.

#### 1. Selection of the monthly mosaic set

The validation metrics concerning the chosen month's association as training input do not lead to a clear dominance of one association over the others. Even if the performance of the models were still improving after 20 epochs, we assumed that the convergence of the models would follow a similar curve between the different associations, as we couldn't assure full training for all associations. One can see in Figure [5.1](#) that the increase in the metrics is the same for all the associations. Results are very close when we look at the extent captured by the groups tested, and it is hard to make a single choice regarding those metrics. However, looking at the boundaries, we can clearly see that two associations stand out from the rest: the May-August-October and the May-August-November. The difference in the MCC Boundary is more than 5%, which suggests the importance of the August month in the boundary identification. The association May-August-November was therefore selected to conduct the following analysis: these results are in line with what is expected. May is the early season image corresponding



to field preparation, while August is the peak of the vegetation and November is the harvest time, providing a good contrast between crops and fields sown at different dates.

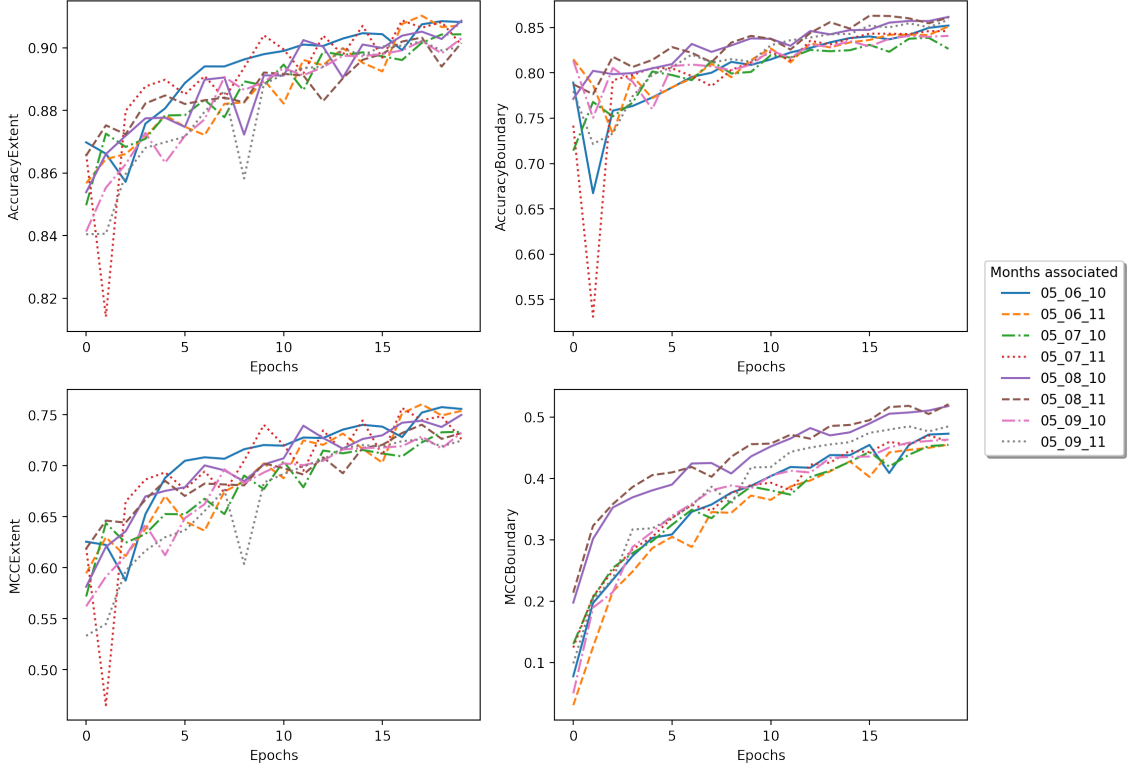


Figure 5.1: Validation metrics comparison obtained with different monthly associations.

## 2. Learning Rate tuning

The first step explores the random variance from training iterations due to initialization. The figure 5.2 illustrates the variability between different training sessions with the same parameters. If we take a look at the variations when the model has fully converged (around 25 epochs here), we see that the variations are greater in extent than the boundaries. The variations concerning the extent are around 2% for the OA and 4% for the MCC, which is two times greater than for the Boundary, where only around 1% is observed for the accuracy and 2% for the MCC. It is therefore important to note that minor differences between two runs with distinct parameters cannot be sufficient to discriminate one from the other.

The second step explores various LR for model training and aims to select the one with the best results. The figure 5.3 shows the result obtained with the three initially tested learning rates. We clearly see that the metrics are way better using the  $10^{-3}$  LR, especially regarding the boundaries. A second manual calibration was tested using various learning rates around  $10^{-3}$  (Figure B.4 in Appendix B). No significant gain is observed using other learning rate values than  $10^{-3}$ , we chose therefore to keep this value for further analysis.

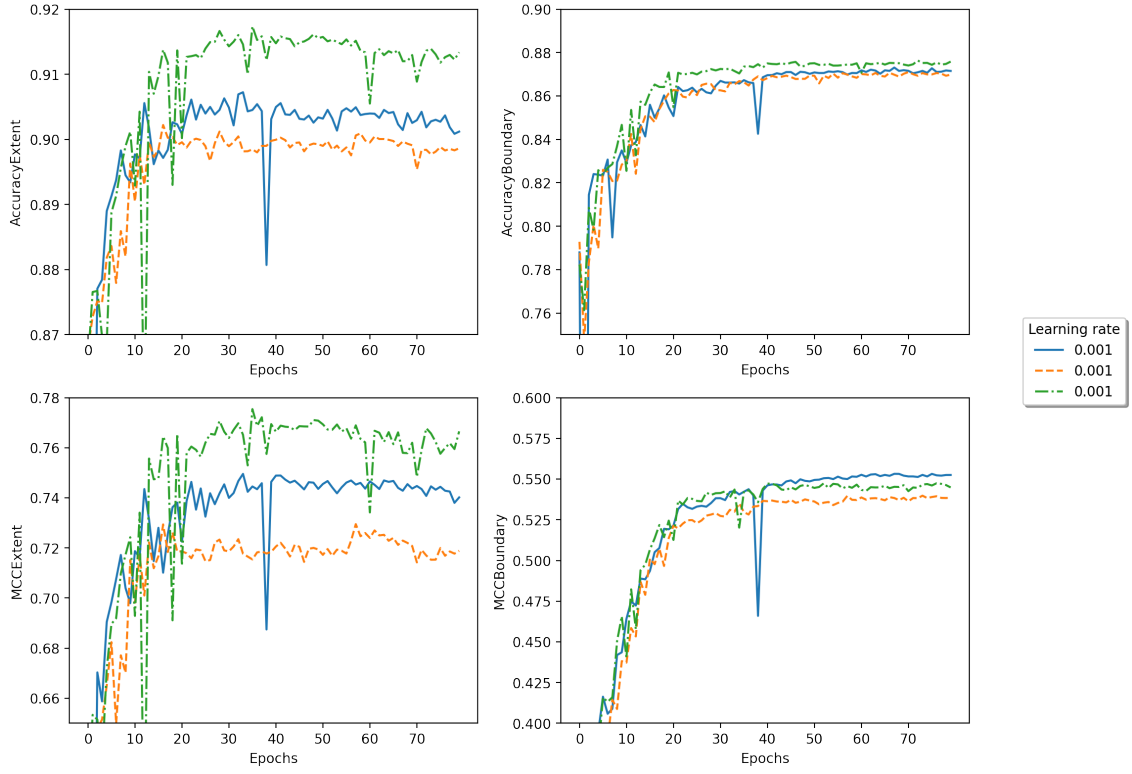


Figure 5.2: Analysis of the variability of the validation metrics using the same LR for every model.

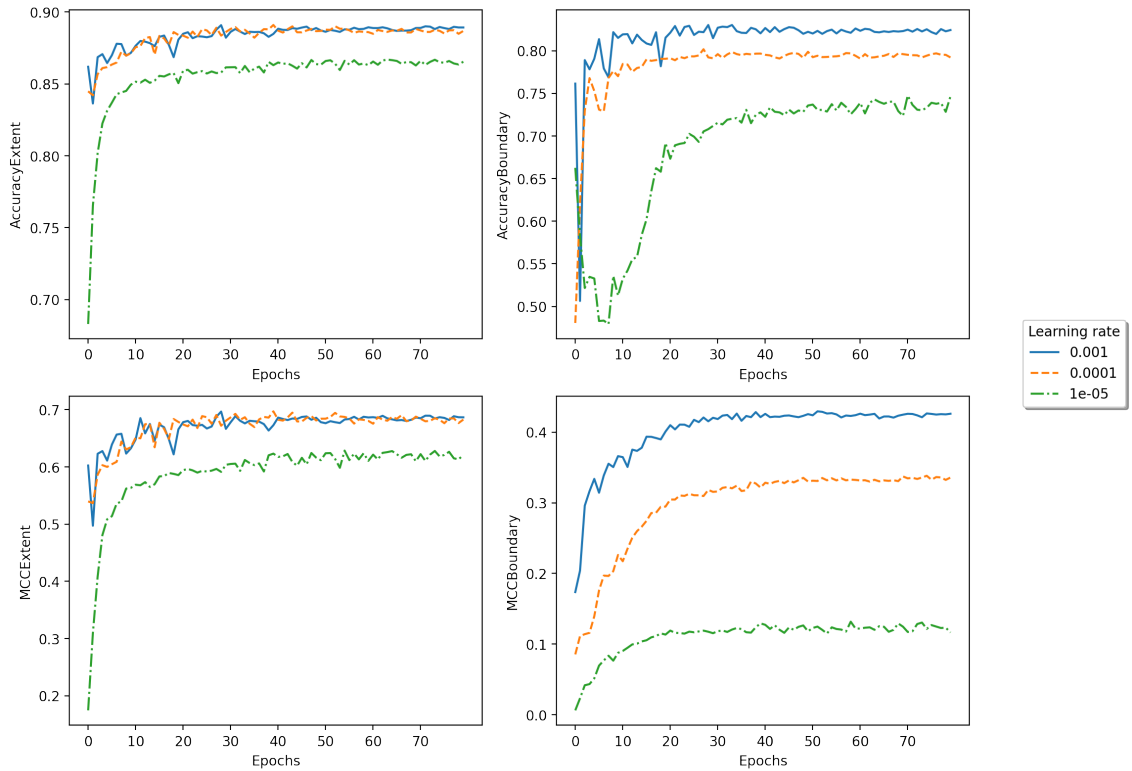


Figure 5.3: Influence of the starting LR on the validation metrics.

The third step looks at the potential benefits of using a decaying learning rate in the training process. We found that adding a scheduler to reduce the learning rate after a certain number of

epochs is not necessarily beneficial for the good training of the model. In this case, the Quality Set responds with an abrupt stop of increase when the learning rate is reduced (figure [B.5](#)).

### 3. Loss Analyses

The Loss Analysis tests different loss compositions during the training and looks for the composition giving the best results, according to the performance metrics and the visual assessment, in the first case. First, the Loss Analysis shows us the clear impact of the HSV layer in the loss calculation on the results. On the one hand, Figure [B.6](#) (Annexe B) shows us the difference in the validation metrics for a training with and without the HSV layer. Curiously, except for the MCC Boundary, it is hard to distinguish any relevant difference between the two runs. Indeed, we saw that differences between two runs with the same parameters can lead to up to 2% differences in accuracies. On the other hand, the visual assessment of the result on Figure [5.4](#) shows a clear difference in the prediction realized on the test set. We clearly see that the HSV layer adds coherence to the predicted field extent area, and that the sides of the fields are more fragmented when working without the HSV layer. Moreover, even within the field extent area, the borders between fields are regular, and more discontinuities are visible on the prediction image without the HSV layer. Therefore, we concluded that the use of the HSV is crucial during the training process of the ResUNet-a model.

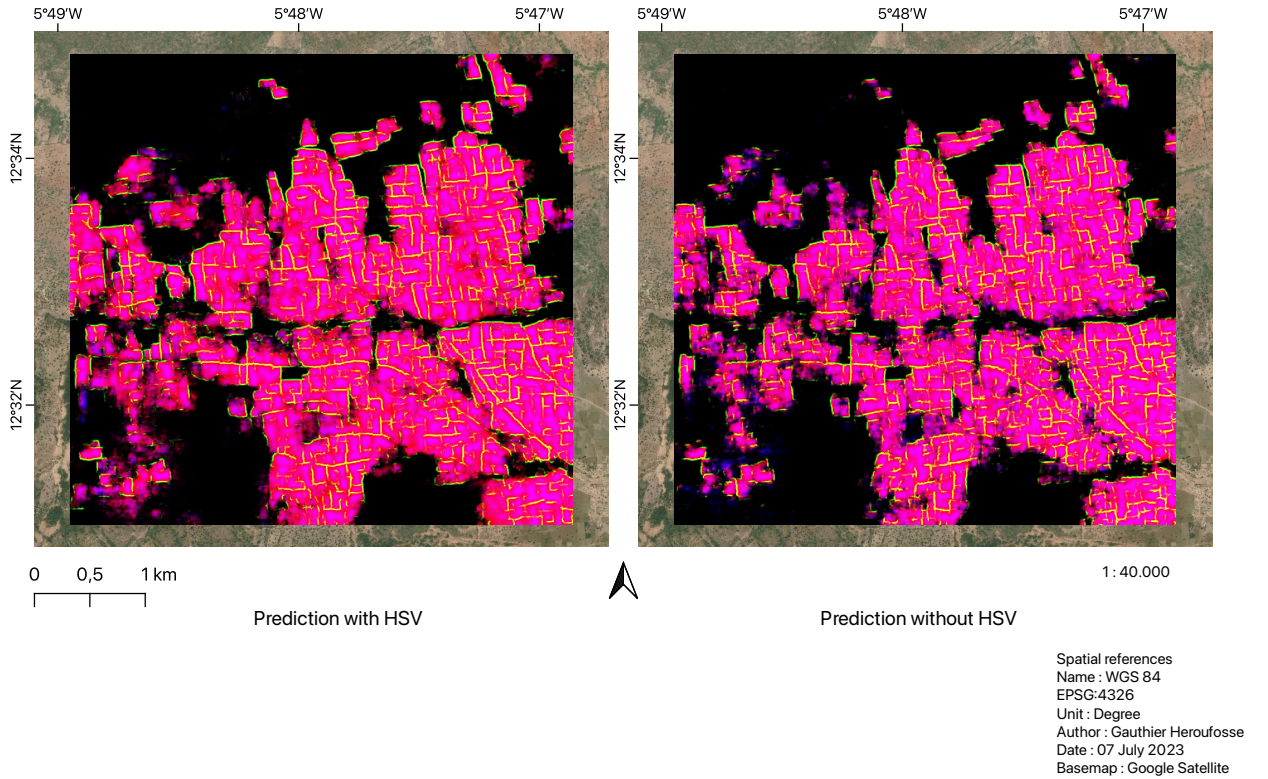


Figure 5.4: Comparison of prediction maps with and without the HSV layer.

The second analysis focused on the distribution of the weights between the losses. As a

reminder, the loss is initially calculated as a balanced average of the four tasks, i.e., Extent mask, Boundary mask, distance to boundary, and HSV layer. Here, the weights were doubled, respectively, for the Extent and the Boundary masks to increase their impact on the training process. Both operations lead to a significant increase in accuracy for their specific tasks. The MCC Extent with the weighted loss by the extent is around 2% bigger than the balanced average loss (see Figure B.7 in Annexe B). A visual comparison was then conducted to assess if one of the other models led to a better prediction result. The figure 5.5 shows both predictions side by side on the Quality Test set, and the figure B.8 in Annexe B shows the reference Quality Test set. At first glance, we can see that the boundaries are much smoother on the prediction with the increased Boundary recognition. Also, boundaries are thinner and more direct. On the other hand, the prediction with increased Extent recognition presents a larger field area, but it also seems more chaotic. A few predicted field pixels are found in the middle of the non-field in the north part of the image and in the central part of the image. We clearly see a separation between the north and south on the image on the right, which is a river. The separation is much more tenuous and less visible on the left image, which doesn't properly reflect the reference. The results were then compared with the initial balanced average loss (visible on Figure 5.4 with HSV), and the prediction using increased Boundary recognition remains the best result.

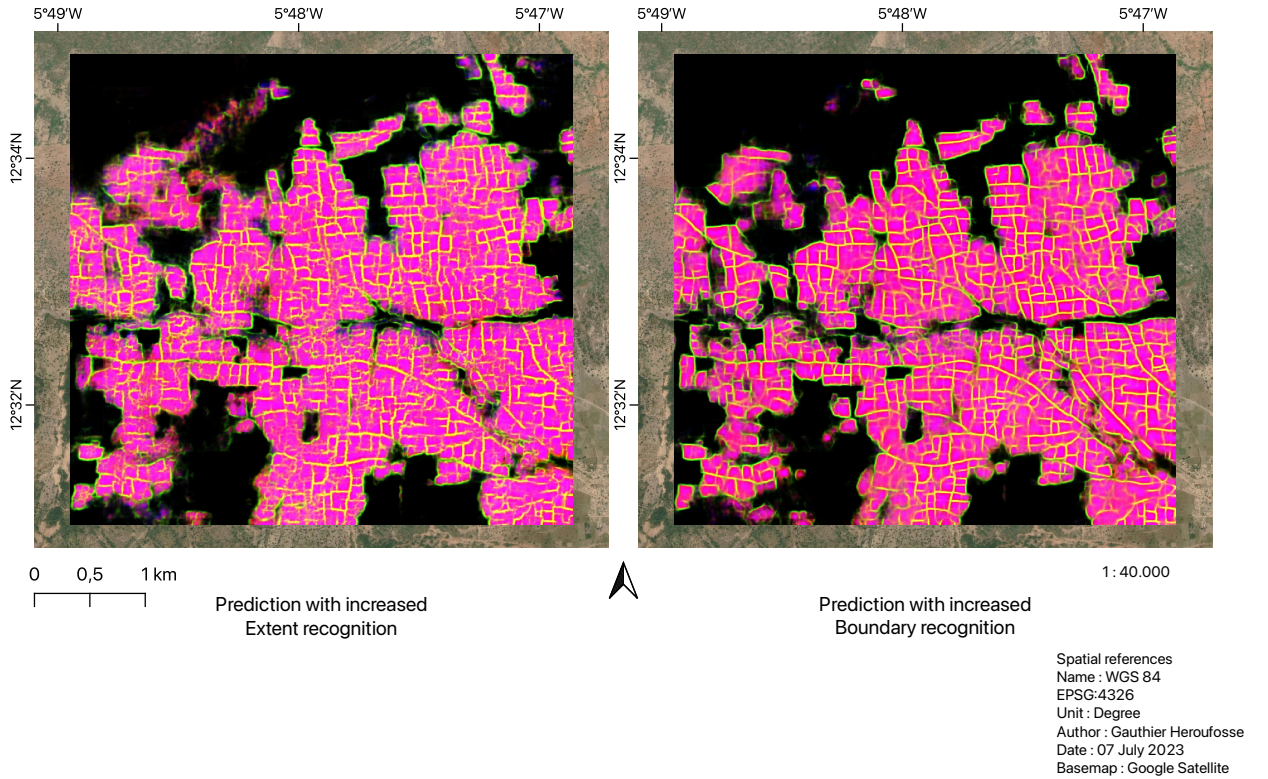


Figure 5.5: Comparison of prediction maps with enhanced recognition of boundaries and extent.

After the calibration analysis, the inputs chosen are the monthly association May-August-

November, and the method is conducted using a fixed learning rate of  $10^{-3}$ , with the HSV layer based on the three NIR bands, and the Tanimoto Loss doubly weighted on the boundary.

### 5.1.2 Assessment of the baseline method

The figure 5.6 presents the validation metrics calculated on the Validation set. The extent related metrics first start to increase, moving from 56% to 62% for the MCC, but after the 10th epoch, the value starts to decrease. This is probably due to an overfitting of the model over the Training set, which the Validation set is independent from. Concerning the boundaries, even if the global accuracy doesn't seem to improve, the MCC value continues to increase after the 10th epoch. The OA is calculated from every correctly classified pixel (binary classification here), and the classes are unbalanced. There are way more non-boundary pixels than boundary ones. Therefore, the accuracy can be quite high, but the result is not that good. The MCC is a robust metric to watch in this situation. The selected model is the one with the highest MCC extent. The boundary threshold parameters of the watershed extraction were empirically tuned to find the best visual result, in accordance with the criteria presented previously (section 4.2.3). Quantitative assessment was then carried out using the post-processed shapefile resulting from the extraction.

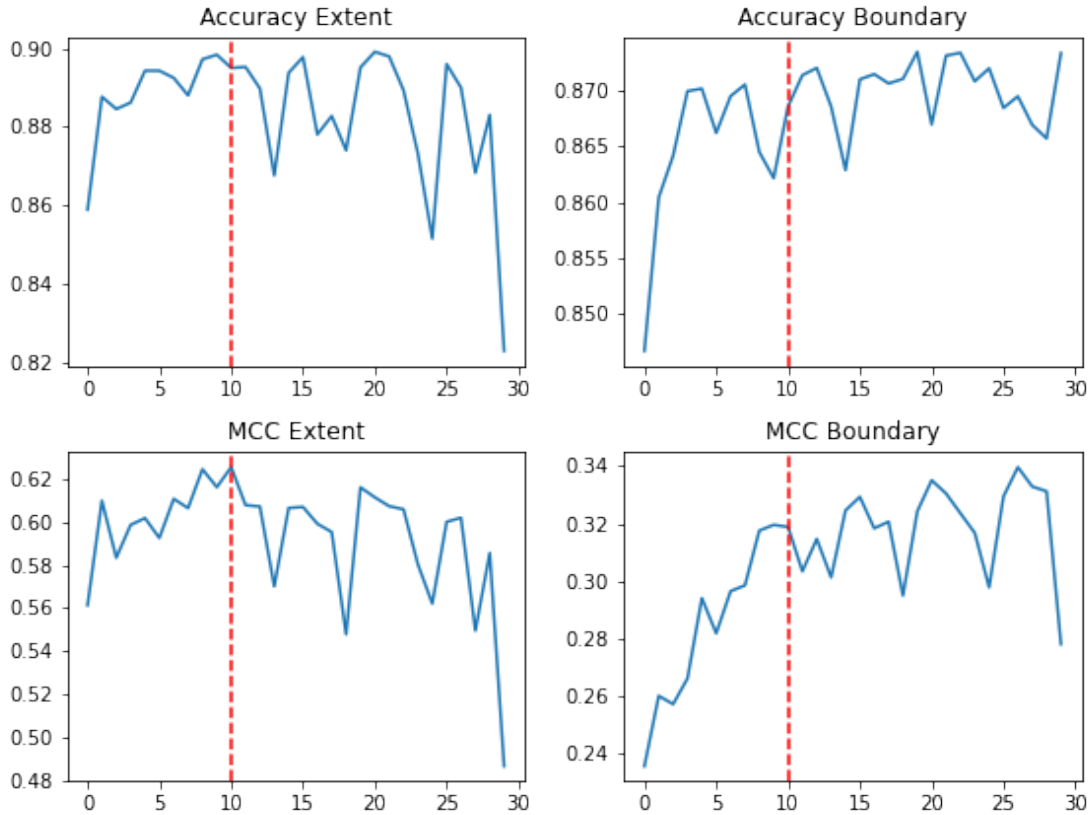


Figure 5.6: Validation metrics for the baseline method. The model chosen is the one with the best result in terms of MCC extent, represented by the vertical red line.

The figure 5.7 presents the final results obtained with the baseline method on the complete



Test set. After the post-processing tasks (see Section 4.2.2.3), 18.185 fields were extracted from the Test set, with an average area of 11 ha. The reference layer gathers 12.023 fields with an average of 15 ha. The Test set is  $40 \times 11 \text{ km}$  in size.

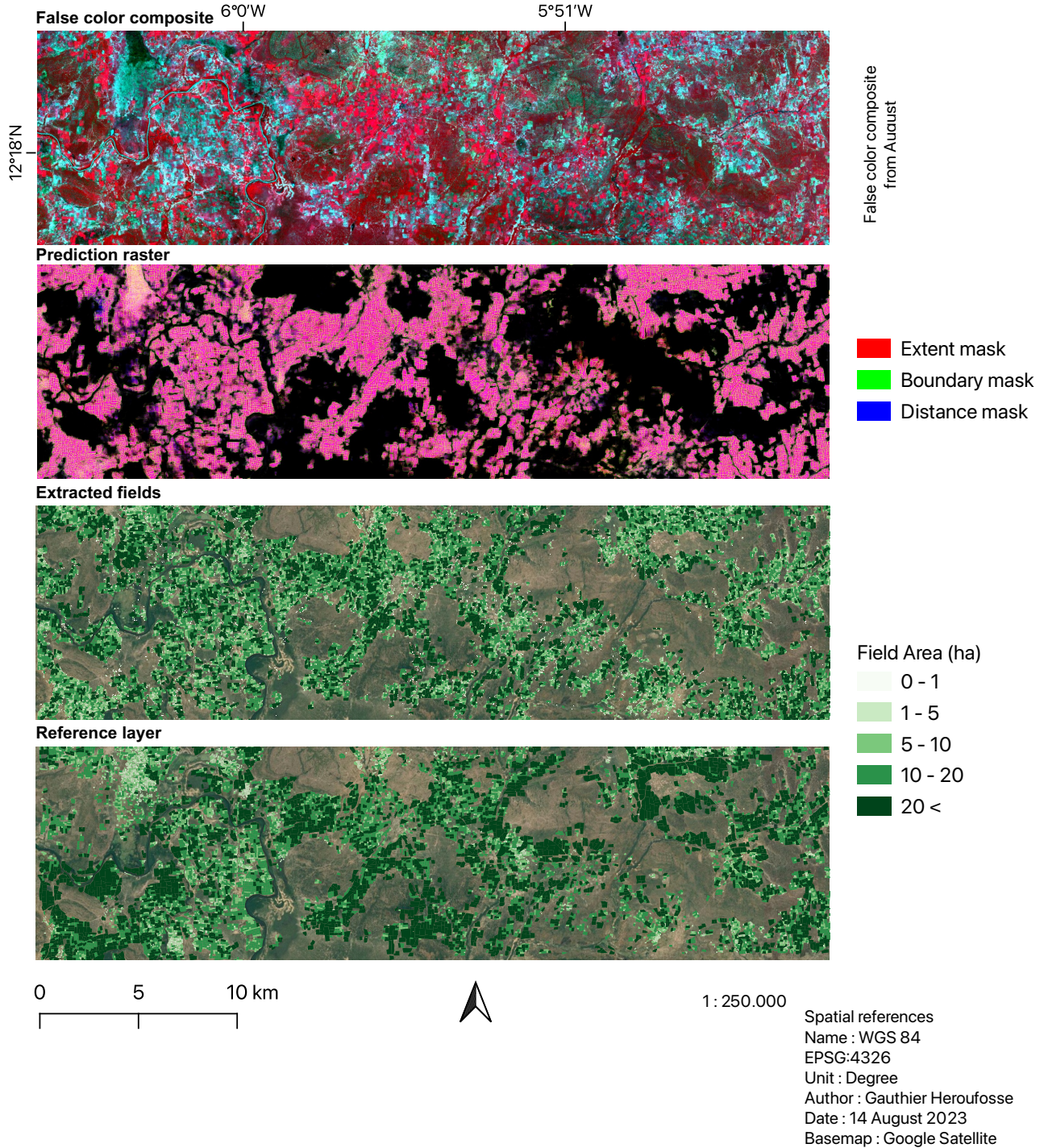


Figure 5.7: Final result obtained with the baseline method. The extracted fields and fields from the reference layer are displayed according to their area. The rose values on the prediction raster correspond to the extent, and the green/yellow to the boundaries.

### 5.1.2.1 Qualitative assessment

The qualitative assessment is made based on the criteria presented in Section 4.2.3. At first sight, no significant difference stands out in the covered area. The field extents seem quite similar on the extracted and reference layers. By going more into details, one can identify some areas where the model finds fields where the reference doesn't indicate so (False Positive). However, after a quick look at the Planet and Google Satellite images, it is easy to see that most of those False Positive are in fact fields, which points out once more the weakness of the reference layer. The most obvious example is at the bottom right of the image, where the reference layer stops 5 fields higher than the extracted layer. On the other hand, some areas are also missed by the model in certain specific environments, : where lots of trees are to be found in the middle of the field plot (Figure 5.8). It is, however, not a major concern as this involves a very small part of the area, and the model still identifies around 60% of the fields.

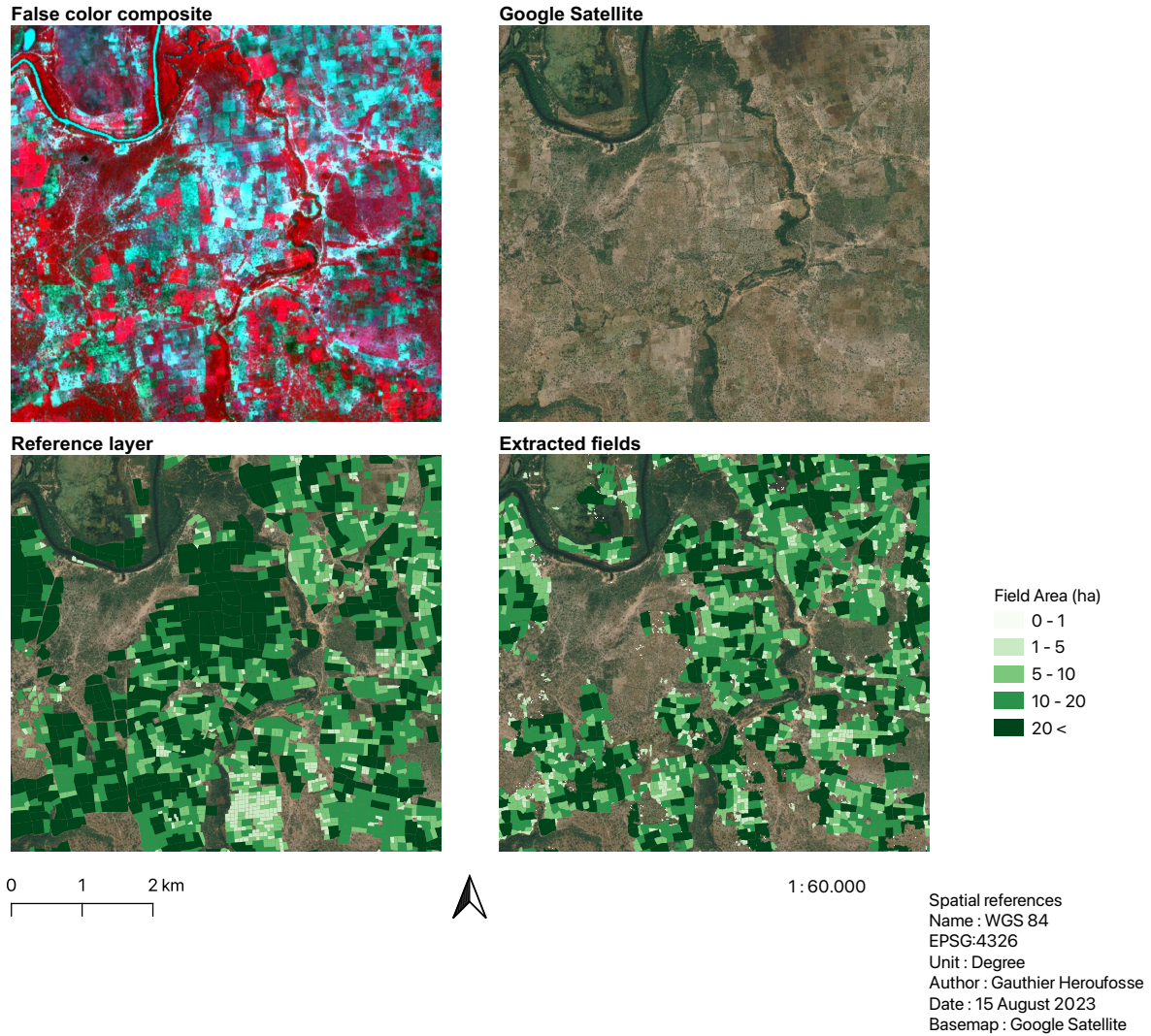


Figure 5.8: Zoom (1) over an area of special interest in the results.

The most notable difference is the field size. One can see in Figure 5.7 that the reference



layer gathers larger fields than the extracted layer. This means that the model is likely to suffer over-segmentation, if we consider the reference layer completely reliable. On Figure 5.9, a wide band of fields in the shape of an "L" is made of fields bigger than  $20ha$  on the reference and mostly of fields smaller than  $20ha$  for the extracted layer. When looking at both satellite images, it is hard to draw a quick conclusion on which segmentation is better. The truth probably lies somewhere between over-segmentation of the extracted fields and under-segmentation of the reference layer.

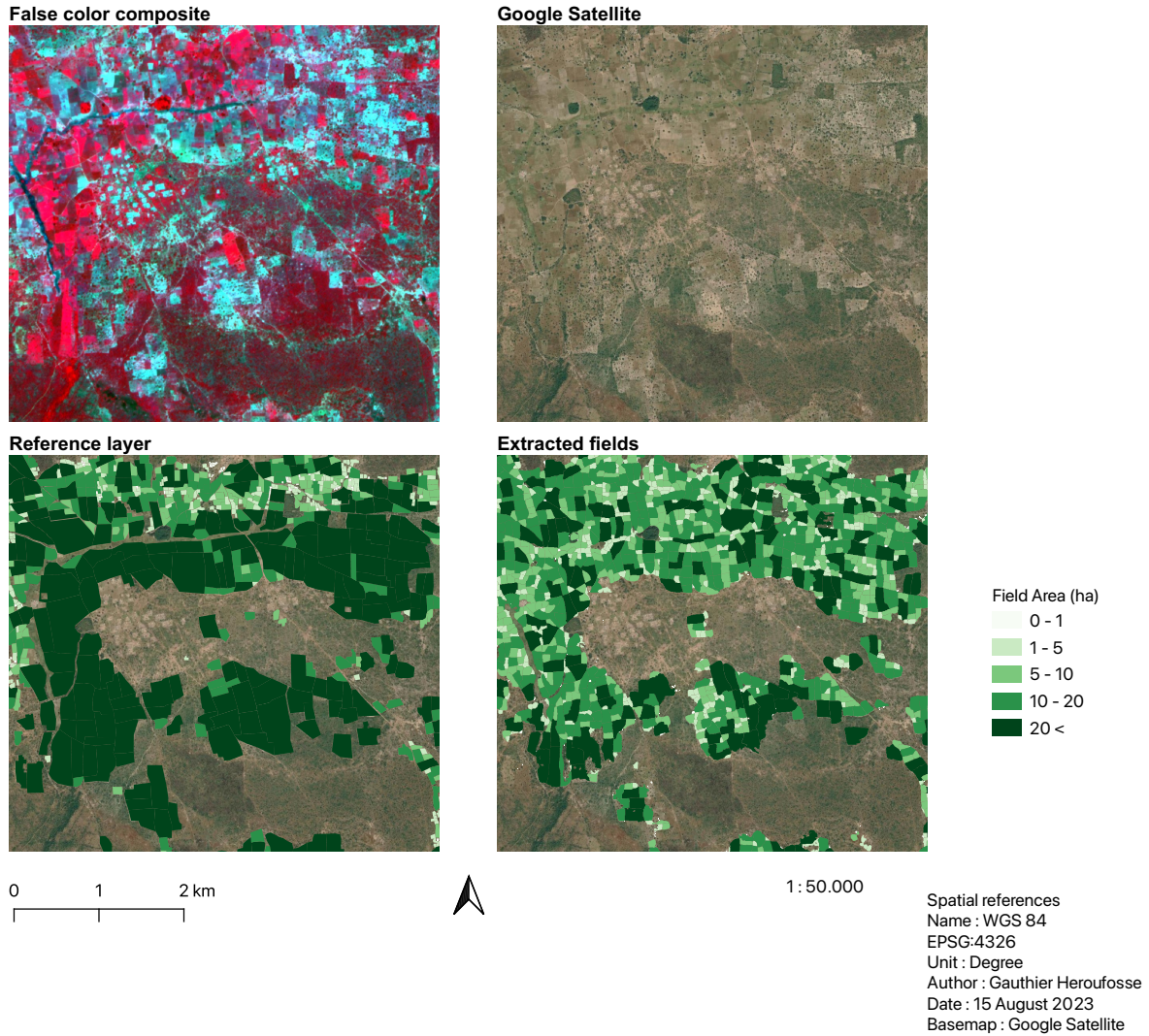


Figure 5.9: Zoom (2) over an area of special interest in the results.

Regarding the field boundaries, it is hard to draw any global conclusion at first sight by viewing the entire testing area. Depending on the land architecture, the boundaries can be very close to the reference layer or, on the other hand, very different from it. This is why the results will be presented with two distinct examples illustrating each of the situations. The first image shows a situation when the extracted boundaries are similar to the manually referenced ones (Figure 5.10). Also, this area has been chosen to have a good-quality reference layer. On

first impression, we clearly recognized the square pattern present on the image on both layers, and it is also confirmed by the two satellite images from Planet and Google. The same lacks previously explained are to be found on this image: when fields are heavily tree charged, the model struggles to consider the field plot as such. The extracted layer recognized pretty well the structures, such as roads and villages, on the image. Concerning the field sizes, we clearly find more small field plots on the extracted image than on the reference layer, which is a sign of over-segmentation. They are usually either located on the edge or in small clusters. We also see at the bottom right of the image some fields in the middle of a bare area, where nothing is cultivated, that can be considered as False Positives. Overall, the impression of this small area (size of  $2.5 \times 2.5 \text{ km}$ ) is good, with some differences not dramatically affecting the quality of the zone.

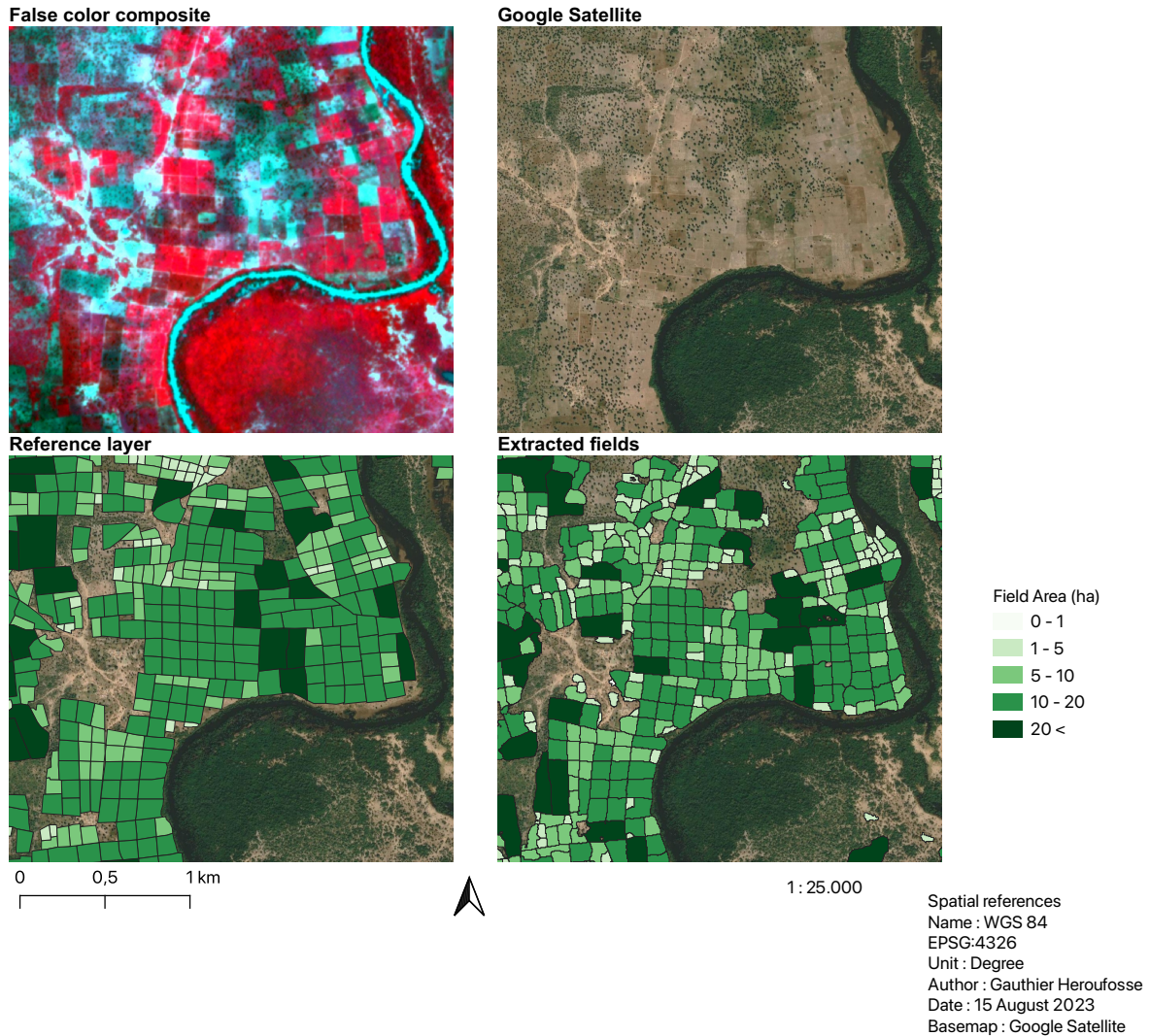


Figure 5.10: Zoom (3) over an area of special interest in the results.

The second image shows us a larger area ( $5 \times 5 \text{ km}$ ) with completely different land management (Figure 5.11). The broad central zone gathers very small field plots with lots of paths



in between. The crop type is also different from the previous image. We see on the reference layer that some fields are missing in the southern part of the zone. Moreover, the False color composite image is also characterized by a large, darker area. It doesn't seem to affect the model that much, as almost all fields are still present. However, the south of the central zone is not well delimited by the model, as the fields are really big and merged together. Around 90% of the extent is still found, but this particular zone suffers from under-segmentation. We cannot consider this area well segmented by the model, as too many errors are to be found. This area is, however, particular, and this is the only place where such a big difference is found in the entire test set.

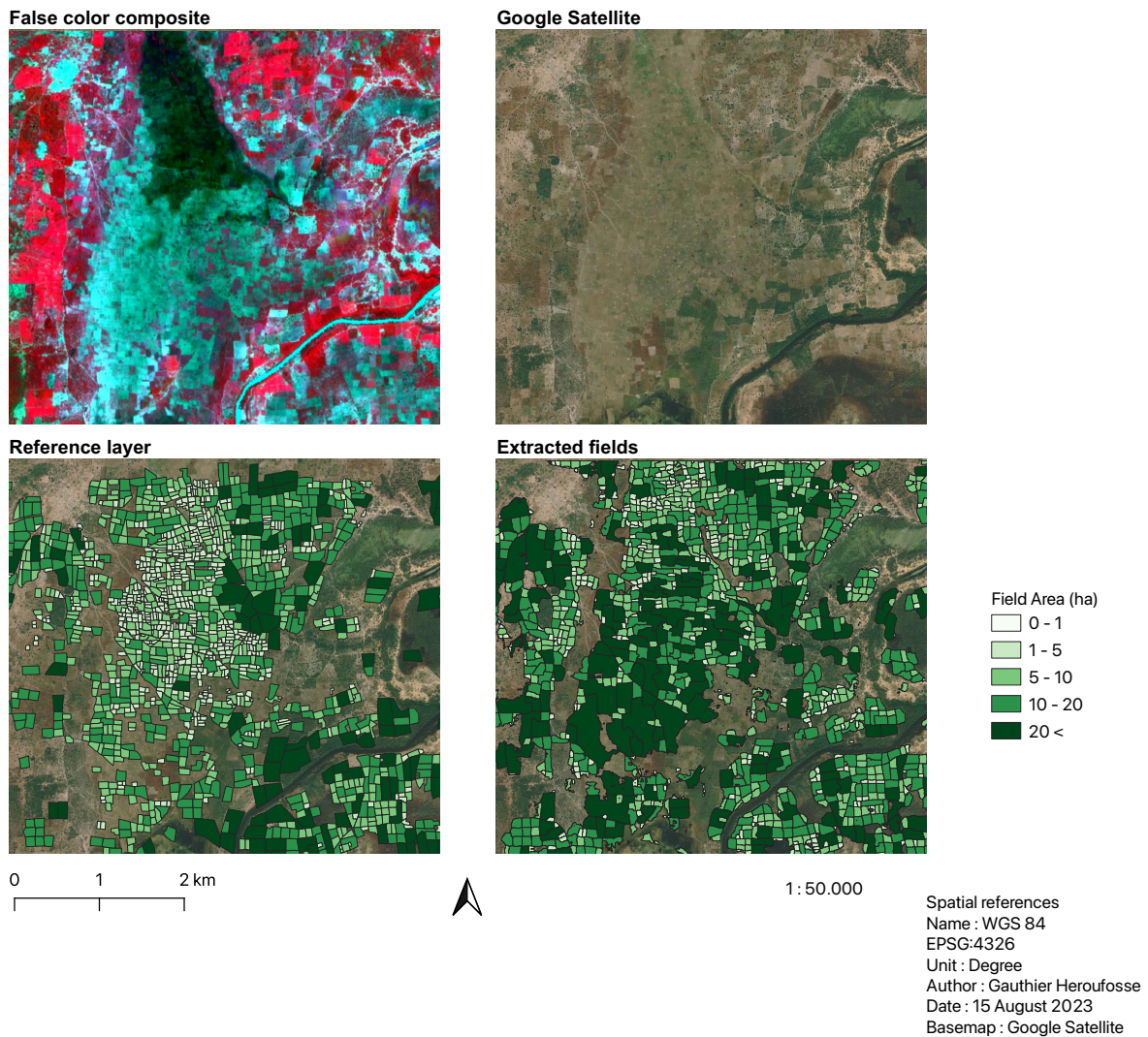


Figure 5.11: Zoom (4) over an area of special interest in the results.

### 5.1.2.2 Quantitative assessment

The table 5.1 presents the final metrics obtained on the entire Test set with the baseline method. With the pixel-based metrics, Extent and Boundaries should be considered independently, while the object-based metrics return a global index for the segmentation. About the Extent, OA is equal to 79.17%. The F1-score and the MCC give, respectively, 79.57% and 0.58, and the Precision and Recall are equal to 77.56% and 81.69%. About the Boundaries, the OA reaches 77.55%, while the F1-score and the MCC are equal to 50.07% and 0.36. The Precision and Recall are equal to 44.62% and 57%. This clearly demonstrates the limitations of the OA when classes are unbalanced, like in the boundary case. The object-based metrics show very interesting results. First, the Precision and Recall are respectively 69.15% and 71.53%, which means that the generated segments have a good overall quality, and that the reference objects are well represented by their corresponding extracted fields. The resulting F-measure<sub>PR</sub> is therefore 70.32%. The RecRate<sub>soft</sub>, which, as a reminder, excludes the under- and over-segmentation errors, gives 86.63%. This goes with a FPR of 40.66%. Those results clearly demonstrate that the model struggles to be completely reliable on the boundary detection. They need to be treated with caution, as the reference layer presents a large number of missing fields (around 10% in estimation), especially at the edge of the images, as presented before. All the reference boundaries are not marked on the reference layer, which also suffers from under-segmentation. Considering the limitations of the data, it is hard to extend the quantitative analysis of the results as the reference is not totally reliable.

Table 5.1: Quantitative assessment of the baseline method using pixel-based and object-based metrics.

| Metrics                     | Global | Fields | Boundaries |
|-----------------------------|--------|--------|------------|
| <i>Pixel-based metrics</i>  |        |        |            |
| Overall Accuracy            | /      | 79.17  | 77.55      |
| F1-score                    | /      | 79.57  | 50.07      |
| MCC                         | /      | 0.58   | 0.36       |
| Precision                   | /      | 77.56  | 44.62      |
| Recall                      | /      | 81.69  | 57.04      |
| <i>Object-based metrics</i> |        |        |            |
| Precision                   | 69.15  | /      | /          |
| Recall                      | 71.53  | /      | /          |
| F-measure <sub>PR</sub>     | 70.32  | /      | /          |
| RecRate <sub>soft</sub>     | 86.63  | /      | /          |
| False Positive Rate         | 40.66  | /      | /          |

## 5.2 Additional analysis

This section brings together the results of the additional analysis carried out after the basic method. The assessment of the Quality set will be conducted first, followed by various analyses that test the results obtained with different types of input data, and explore the generalization capacities of the chosen method.

### 5.2.1 Quality Test Set assessment

This analysis explores the benefits of having access to a high-quality reference layer by comparing the respective results obtained with the Quality set and its corresponding non-modified set (called Original set hereafter). Two models were trained for 200 epochs with the optimal parameters retrieved from the baseline method. The Test layer was also improved, like the Quality set, with rigorous control over the extent and boundaries of the fields present in the area. The table 5.2 shows the final results obtained with both sets.

Table 5.2: Quantitative assessment of the Quality Test set using pixel-based and object-based metrics.

| Metrics                     | Quality Set |        |            | Original Set |        |            |
|-----------------------------|-------------|--------|------------|--------------|--------|------------|
|                             | Global      | Fields | Boundaries | Global       | Fields | Boundaries |
| <i>Pixel-based metrics</i>  |             |        |            |              |        |            |
| Overall Accuracy            | /           | 87.92  | 76.52      | /            | 87.92  | 74.42      |
| F1-score                    | /           | 90.13  | 49.92      | /            | 89.82  | 48.76      |
| MCC                         | /           | 0.76   | 0.35       | /            | 0.75   | 0.32       |
| Precision                   | /           | 85.22  | 48.85      | /            | 87.35  | 46.74      |
| Recall                      | /           | 95.64  | 51.04      | /            | 92.44  | 50.78      |
| <i>Object-based metrics</i> |             |        |            |              |        |            |
| Precision                   | 49.12       | /      | /          | 47.81        | /      | /          |
| Recall                      | 80.54       | /      | /          | 76.82        | /      | /          |
| F-measure                   | 61.02       | /      | /          | 58.93        | /      | /          |
| RecRate soft                | 95.39       | /      | /          | 93.02        | /      | /          |
| False Positive Rate         | 16.42       | /      | /          | 14.92        | /      | /          |

When looking at the pixel-based metrics, results are surprisingly close, especially about the field extent. Indeed, the MCC varies only by 0.01, and the other field-related metrics have the same behavior. The pixel-based Recall is very high for both the Quality and Original sets, which indicates that both sets recognized pretty well the extent of this region. For the pixel-based Precision, we see that the value is even higher for the Original set than for the Quality set. This goes along with the FPR also lower for the Original set. About the boundaries, we see a bit more variation, with the Quality set being obviously more accurate than the Original set. Obviously, Overall Accuracy is still not relevant to interpret in the context of Boundaries. For the MCC, a variation from 0.32 to 0.35 is observed when correcting the Original set, which is not very much, and the same can be said of the F1-score. This analysis does not allow us to conclude

that the changes made to the Original set to correct it brought a clear improvement. Even if the boundaries seem to better reflect reality, this goes with a surprising decrease in extent quality. That is quite an unexpected result, because the original set can not be considered confidence data in the training of a model. This needs to be relativized, however, as the Quality set is only 1.903 fields large, and although the validation is close to it, some differences are still found between the two areas. For instance, the validation area includes more wooded zones. They have been identified as areas of concern in the assessment of the baseline method (section 5.1.2).

### 5.2.2 Very High Resolution images

The second additional analysis tries to use VHR images (1m spatial resolution) for the training of a ResUNet-a model. The figure 5.12 displays the results in the validation area. One can see at first glance that the obtained results are not exploitable. The predicted raster is made up of an inconsistent extent prediction, and the boundaries are not even visible on the edge of the field extent.

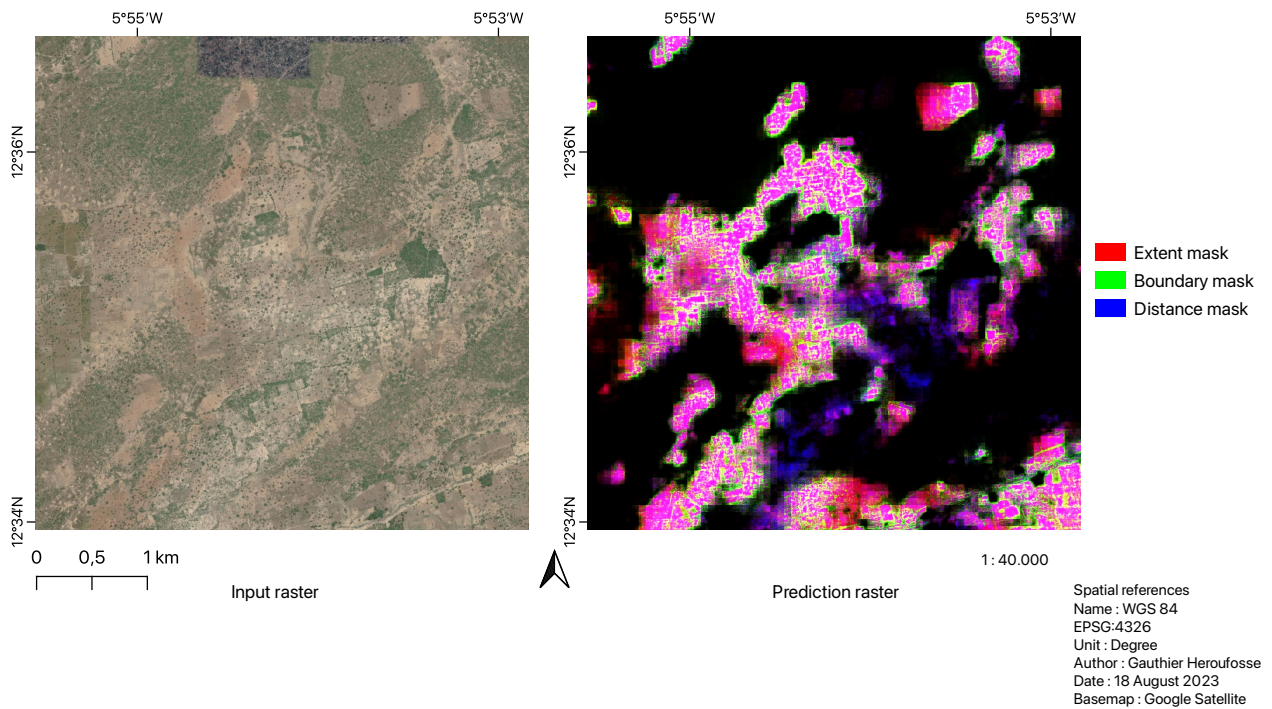


Figure 5.12: Prediction raster from the testing of a model trained with VHR Maxar images. The rose values correspond to the extent, and the green/yellow to the boundaries.

When zooming in at field size, the issues can be better identified. First, the visible squares correspond to the inference method using sliding windows to make the prediction. The visual impact is much greater than with the basic method. This is probably due to the fact that many of the squares have no fields at all, which is less the case with Planet’s images, since the resolution is five times greater. In addition, the trees in the middle of the plots are all



singularly identified as boundaries, resulting in small areas included in the actual plot. This problem could be solved by an appropriate post-processing method that removes boundaries of a certain shape and size. The model also still struggles to correctly identify the extent. Some bare areas or even roads are falsely classified as fields. This could probably be solved by increasing the training of the model. In conclusion, a separate assessment of the ability of VHR images to perform field segmentation at large scales should be carried out.

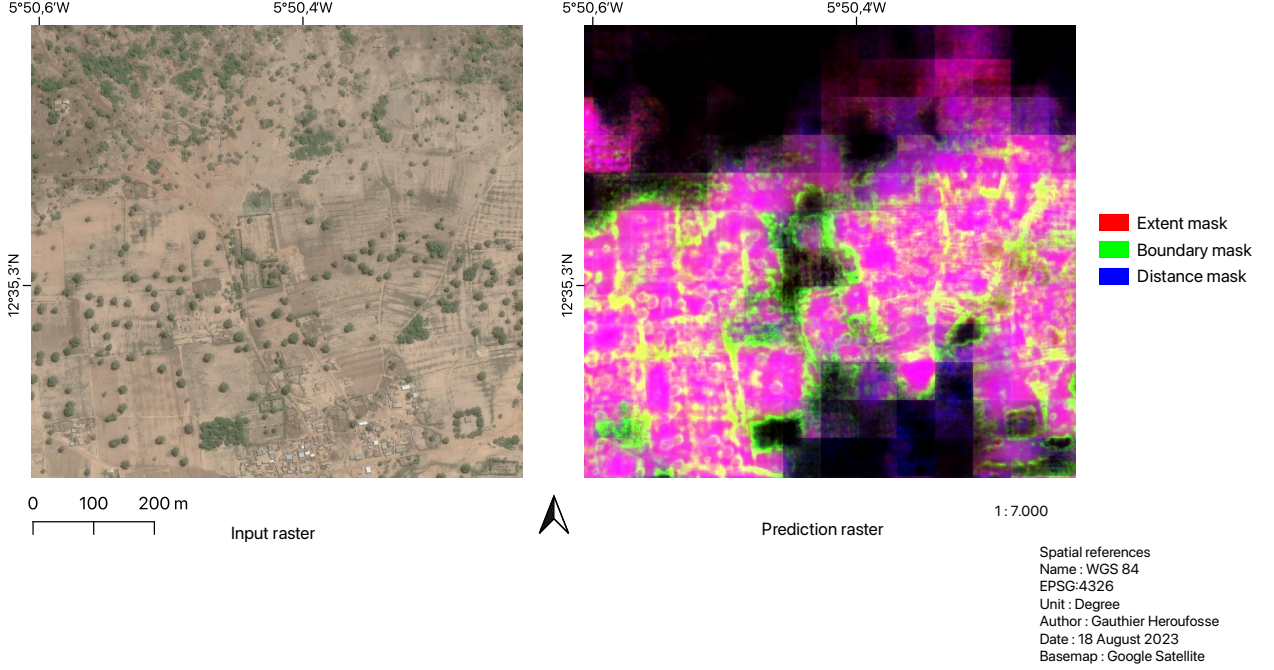


Figure 5.13: Zoom over an area of special interest in the results from the VHR images

### 5.2.3 Model portability within Africa

The model's portability across space was estimated with images from the second ROI (see [4.1.1](#)), which is located 130km south of the main ROI. The result was obtained using the best model resulting from the baseline method.

The figure [5.14](#) presents the prediction realized over the entire ROI. The first impression the extracted fields layer gives is that it is sparser than the reference layer. Indeed, especially in the center of the ROI, lots of fields are missing, where the north and west parts of the image seem to correspond quite well. After a deeper analysis, it was found that one of the monthly Planet mosaics was not in good shape (November). The values were corrupted at some locations, giving those rose values on the False Color composite visible on figure [5.15](#). The quality assessment will therefore be conducted only on the first north half of the result. Most of the extent is shared with the reference layer, and no major difference is to be found. The number of False Positive doesn't seem high, with the majority of the extracted fields also being referenced.

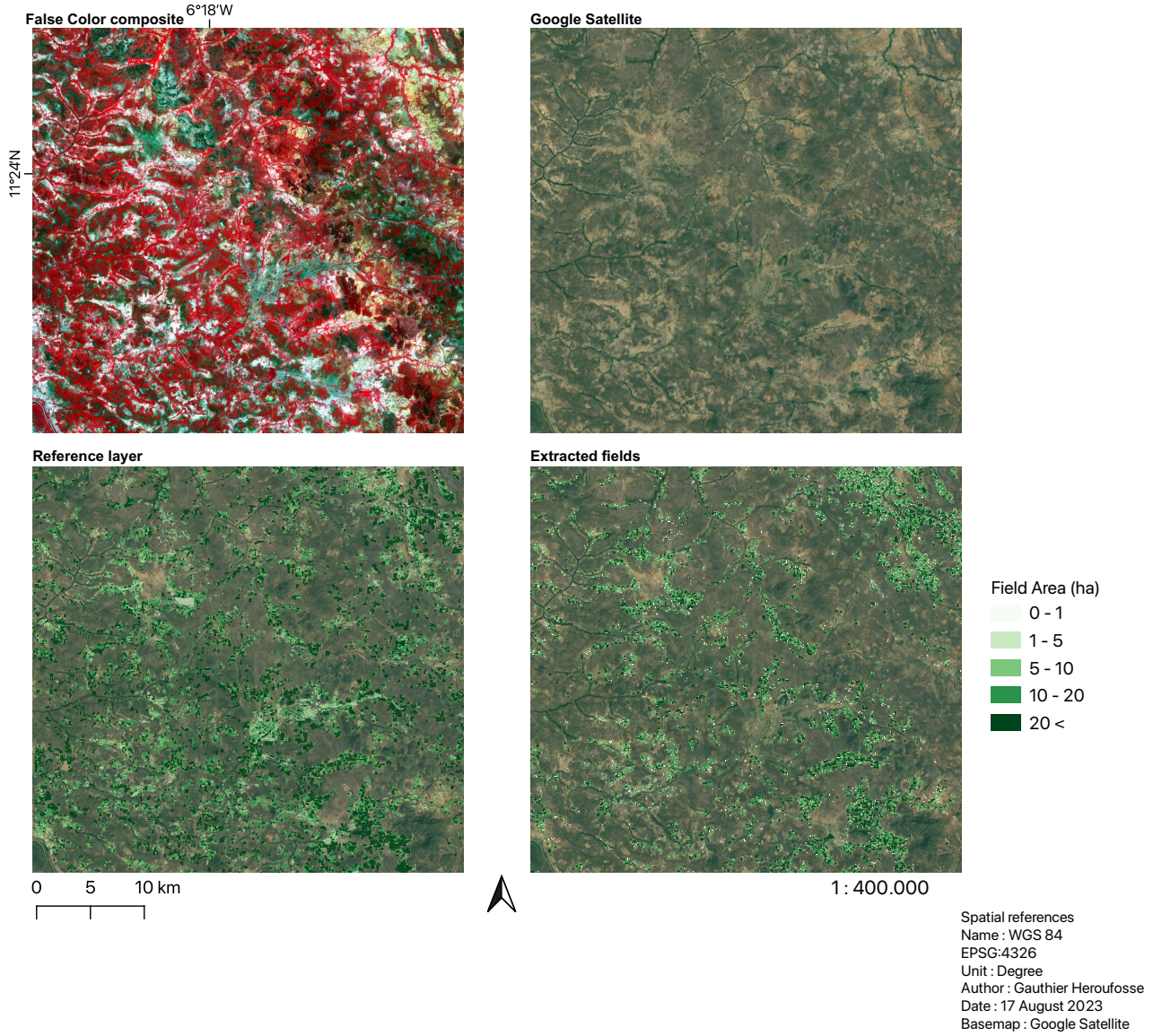


Figure 5.14: Map of the results over the Sikasso region. The extracted fields and fields from the reference are displayed according to their area.

However, the completeness of the result is not great. In each area where fields are found, more are marked on the reference and can be seen using the Planet and Google satellite images. About the boundaries, the large elements (roads, villages, etc.) are found by the model, and the limits are well respected. Inside the field extent, plots are not properly delimited, and fields can't be considered correctly segmented.

Looking at the quantitative metrics, the pixel metrics are in line with the visual assessment : the MCC extent is 0.56, and 0.31 for the boundaries. But the field Recall is only 48.69%, which supports the visual assessment about the completeness. The Precision is on the other side 78.27%, which suggests that found fields are mostly also present on the reference layer. In this case, the OAs are not relevant at all, as the area is less densely populated with field



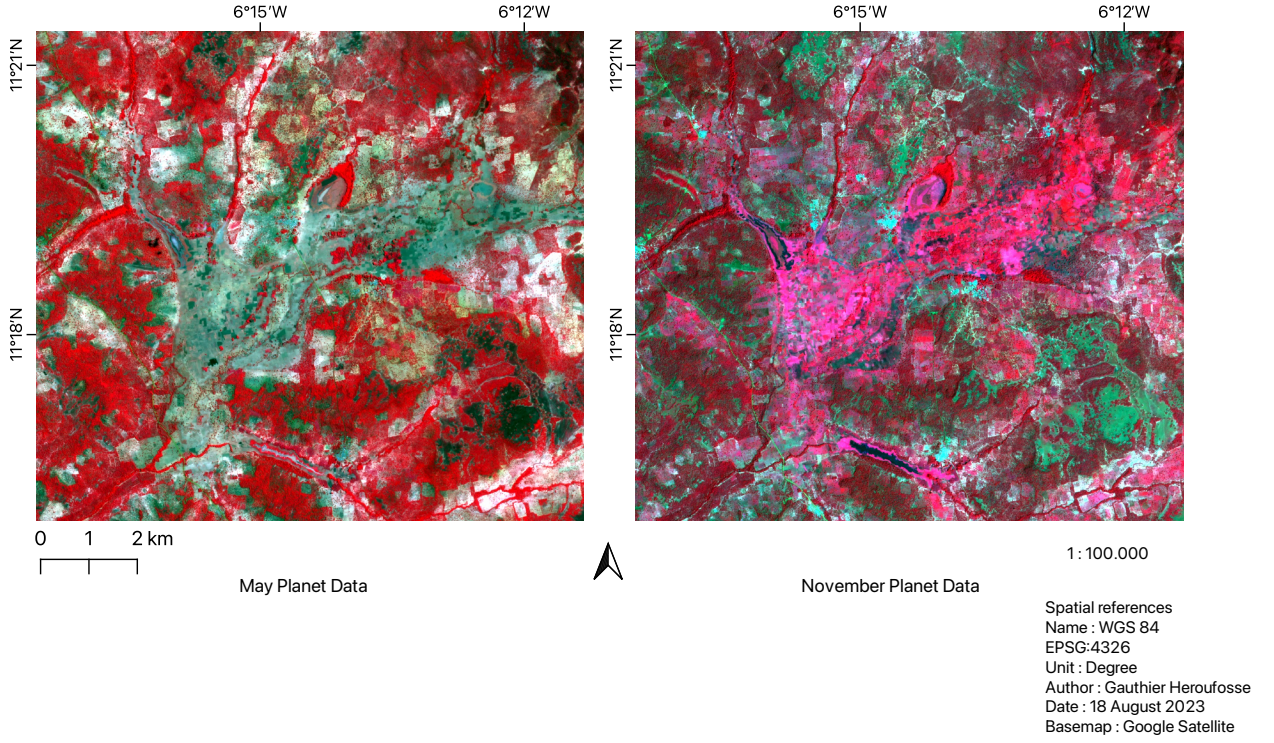


Figure 5.15: Example of data corruption over the Sikasso region. The False Color composite from November shows unusual values.

plots (see [4.1.1](#)). Lots of pixels are correctly classified as "non-field" (or "non-boudary"), which draws the metric really high. About the object-based metrics, the  $\text{RecRate}_{\text{soft}}$  is only 58.76%, which is also in line with the completeness argument. The false positive rate is nevertheless high at 23.78%. Precision and Recall indicate that found fields are mostly in line with their corresponding reference objects, and the  $\text{F-measure}_{\text{PR}}$  gives us their combined score for comparison with further studies.

### 5.3 Discussion and perspectives

The results show that there is still room for improvement in the large-scale field extraction process in the context of smallholder agriculture. The baseline method gives us a good approximation of the agricultural land segmentation but cannot be considered completely reliable. Especially regarding the field extent, the results were promising, as no major difference was found between the reference and the predicted layer. Moreover, the visual assessment confirms the accurate representation of reality by the field extent and mitigates the difference obtained with the reference layer. However, the boundary predictions were not really on point, showing drastic differences with reality in some large areas of the prediction zone. Nevertheless, some areas were particularly well delimited by the model, showing some encouraging signs. As a

Table 5.3: Quantitative assessment of the Sikasso region using pixel-based and object-based metrics.

| Metrics                     | Global | Fields | Boundaries |
|-----------------------------|--------|--------|------------|
| <i>Pixel-based metrics</i>  |        |        |            |
| Overall Accuracy            | /      | 88.51  | 90.33      |
| F1-score                    | /      | 60.03  | 35.94      |
| MCC                         | /      | 0.56   | 0.31       |
| Precision                   | /      | 78.27  | 41.89      |
| Recall                      | /      | 48.69  | 31.47      |
| <i>Object-based metrics</i> |        |        |            |
| Precision                   | 62.55  | /      | /          |
| Recall                      | 67.32  | /      | /          |
| F-measure <sub>PR</sub>     | 64.85  | /      | /          |
| RecRate <sub>soft</sub>     | 58.76  | /      | /          |
| False Positive Rate         | 23.78  | /      | /          |

reminder, the reasonably reachable target for this project was set to be a dataset with a good prediction with respect to field location, but not too severe regarding boundary accuracy. This is roughly the description of what was achieved with the baseline method of the master’s thesis.

The secondary objective was to explore the generalization capacities of the methodology through the use of better input data quality, using VHR images, and assess the model’s portability over a different region. The additional analyses gave mixed results. The benefit of using better reference data couldn’t really be proved in this specific study, which asks for further examinations. The use of VHR images with the ResUNet-a model was not a success. The model struggles to identify both the field extent and the boundaries, and a specific assessment of the ability of VHR to perform field segmentation at a large scale should be conducted. The portability of the model within Africa gives good results, but suffers from a lack of correct input data.

Several sources of error have been identified. The most important drawback is the lack of confidence in the input data. Without an appropriate reference dataset, the DL model will classify features incorrectly and quickly overfit to match corrupted labels due to its ability to learn any complex function. This reduces the quality of the prediction and affects the metrics. In addition, the reference test set is also unreliable, resulting in inaccurate measurements. Qualitative assessment is therefore the best way to validate the results in this case. Large-scale application means that the chosen ROI is more likely to be composed of several landscapes, each with its associated land-use reflecting value. This adds complexity to the input data and makes model training harder. In this case, high-quality data is even more necessary. This is directly related to another source of error that comes from the calibration analysis of the Quality set. The Quality set is only about 3% in area of the quality controlled set. Although it was chosen because of its heterogeneity in the landscape, it cannot realistically be considered

representative of the whole dataset. The optimized parameters found with the Quality set are therefore unlikely to be the same as the optimal parameters for the complete dataset. A quick assessment of the optimal parameters applied to the reference dataset was, however, done in order to check if the overall response of both sets was the same, which was the case. Another source of error is the use of more than four bands in the training of the ResUNet-a model. The model was initially built to receive a 4-band input (RGB and NIR), and the HSV layer is supposed to be made out of the three RGB layers. Here, the HSV layer is built using the three NIR layers from each of the used months, which causes it to take a different value from what it is supposed to take. This changes the weight associated with this output in the loss calculation.

A few recommendations can be made to improve the methodology. In the input image, it would be useful to measure the influence of the blue band in the training process. The blue band has low discriminating power when speaking about different vegetation species (USGS, 2023). With the soil being mainly covered by vegetation species in the SSA region, it might be beneficial to remove it. The results obtained using a spectral index such as NDVI could also be assessed, as it requires less computational power thanks to the combination of the information of two major bands (red and NIR) within one band. The first obtained result could also be used to update the reference layer in a sort of self-validation loop in the absence of certainty about the reference data. The automation of the post-processing method to select the threshold giving the best metrics could also be a slight improvement to the project.

Finally, it would be necessary to compare the results with other state-of-the-art modelling approaches. The models from the related studies presented in section 2.7 could be used on the same area to check whether ResUNet-a performs best in the SSA.

## Chapter 6: Conclusion

This thesis aimed to evaluate the performance of an automatic field extraction methodology from High Resolution satellite images in the context of smallholder agriculture. The methodology relies on a Deep Learning model trained with satellite images from 2021 captured by the Planet satellite and the reference dataset. The first step consisted of a calibration analysis to set the best hyperparameters for the model training. The baseline method was then evaluated with carefully selected metrics showing the performance of our model. The last step includes the assessment of the model's portability within Africa, as well as an exploration of possible improvements by adjusting the quality of the input data and the imagery's spatial resolution.

Overall, the results were mixed. The model is mostly capable of finding the fields extent, but struggles to capture field boundaries correctly. Nevertheless, some areas were particularly well delimited by the model, showing a promising improvement perspective with high quality data. The portability of the model is decent, but it suffers from land cover and topography differences. The use of Very High Resolution images was not a success, as the model is not capable of correctly identifying fields extent and boundaries. Major drawbacks are the lack of reliability of the reference dataset and the complexity of the agricultural arrangement, with the field area being on average 1.2 ha.

Finally, recommendations and improvement opportunities were discussed for a potential continuation of experiments. Comparing the results with other state-of-the-art modeling approaches would also be beneficial.



# Bibliography

- Abramova, I. O. 2022. “The Population of Africa under the Conditions of Transformation of the World Order.” *Herald of the Russian Academy of Sciences* 92 (Suppl 14): S1306–S1315. ISSN: 1019-3316, accessed August 19, 2023. <https://doi.org/10.1134/S1019331622200023>. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10037377/>.
- Badrinarayanan, Vijay, Alex Kendall, and Roberto Cipolla. 2017. “SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation.” Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39, no. 12 (December): 2481–2495. ISSN: 1939-3539. <https://doi.org/10.1109/TPAMI.2016.2644615>.
- Belgiu, Mariana, and Ovidiu Csillik. 2018. “Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis” [in en]. *Remote Sensing of Environment* 204 (January): 509–523. ISSN: 0034-4257, accessed November 25, 2022. <https://doi.org/10.1016/j.rse.2017.10.005>. <https://www.sciencedirect.com/science/article/pii/S0034425717304686>.
- Bengio, Yoshua. 2012. *Practical recommendations for gradient-based training of deep architectures*. ArXiv:1206.5533 [cs], September. Accessed July 6, 2023. <https://doi.org/10.48550/arXiv.1206.5533>. <http://arxiv.org/abs/1206.5533>.
- Blaschke, T. 2010. “Object based image analysis for remote sensing” [in en]. *ISPRS Journal of Photogrammetry and Remote Sensing* 65, no. 1 (January): 2–16. ISSN: 0924-2716, accessed October 19, 2022. <https://doi.org/10.1016/j.isprsjprs.2009.06.004>. <https://www.sciencedirect.com/science/article/pii/S0924271609000884>.
- Blaschke, Thomas, Geoffrey J. Hay, Maggi Kelly, Stefan Lang, Peter Hofmann, Elisabeth Addink, Raul Queiroz Feitosa, et al. 2014. “Geographic Object-Based Image Analysis – Towards a new paradigm” [in en]. *ISPRS Journal of Photogrammetry and Remote Sensing* 87 (January): 180–191. ISSN: 0924-2716, accessed November 19, 2022. <https://doi.org/10.1016/j.isprsjprs.2013.09.014>. <https://www.sciencedirect.com/science/article/pii/S0924271613002220>.

- Boehner, J., Thomas Selige, and Andre Ringeler. 2006. "Spatial pattern recognition by means of representativeness measures" [in en], 1:110–112. Hamburg, Germany: IEEE. ISBN: 978-0-7803-5207-0, accessed November 20, 2022. <https://doi.org/10.1109/IGARSS.1999.773417>. <http://ieeexplore.ieee.org/document/773417/>.
- Canny, John. 1986. "A Computational Approach to Edge Detection." Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence, *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-8, no. 6 (November): 679–698. ISSN: 1939-3539. <https://doi.org/10.1109/TPAMI.1986.4767851>.
- Castillejo-González, Isabel Luisa, Francisca López-Granados, Alfonso García-Ferrer, José Manuel Peña-Barragán, Montserrat Jurado-Expósito, Manuel Sánchez de la Orden, and María González-Audicana. 2009. "Object- and pixel-based analysis for mapping crops and their agro-environmental associated measures using QuickBird imagery" [in en]. *Computers and Electronics in Agriculture* 68, no. 2 (October): 207–215. ISSN: 0168-1699, accessed November 25, 2022. <https://doi.org/10.1016/j.compag.2009.06.004>. <https://www.sciencedirect.com/science/article/pii/S0168169909000982>.
- Chauvin, Richard Durbin. 1995. "Backpropagation: The Basic Theory." In *Backpropagation*. Num Pages: 34. Psychology Press. ISBN: 978-0-203-76324-7.
- Chen, Dongping Ming, Lu Zhao, Beiru Lv, Keqi Zhou, and Yuanzhao Qing. 2018. "Review on High Spatial Resolution Remote Sensing Image Segmentation Evaluation." *Photogrammetric Engineering & Remote Sensing* 84, no. 10 (October): 629–646. <https://doi.org/10.14358/PERS.84.10.629>.
- Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L. Yuille. 2016. *Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs*. ArXiv:1412.7062 [cs], June. Accessed May 15, 2023. <https://doi.org/10.48550/arXiv.1412.7062>. <http://arxiv.org/abs/1412.7062>.
- . 2018. "DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs." Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 40, no. 4 (April): 834–848. ISSN: 1939-3539. <https://doi.org/10.1109/TPAMI.2017.2699184>.
- Chen, George Papandreou, Florian Schroff, and Hartwig Adam. 2017. *Rethinking Atrous Convolution for Semantic Image Segmentation*. ArXiv:1706.05587 [cs], December. Accessed October 14, 2022. <https://doi.org/10.48550/arXiv.1706.05587>. <http://arxiv.org/abs/1706.05587>.

- Chicco, Davide, and Giuseppe Jurman. 2020. “The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation.” *BMC Genomics* 21, no. 1 (January): 6. ISSN: 1471-2164, accessed July 22, 2023. <https://doi.org/10.1186/s12864-019-6413-7>. <https://doi.org/10.1186/s12864-019-6413-7>.
- Climate data for cities worldwide - Climate-Data.org*. 2023. Accessed June 27, 2023. <https://en.climate-data.org/>.
- Costa, Hugo, Giles M. Foody, and Doreen S. Boyd. 2018. “Supervised methods of image segmentation accuracy assessment in land cover mapping” [in en]. *Remote Sensing of Environment* 205 (February): 338–351. ISSN: 0034-4257, accessed May 16, 2023. <https://doi.org/10.1016/j.rse.2017.11.024>. <https://www.sciencedirect.com/science/article/pii/S0034425717305734>.
- Dai, Wen, Jiaming Na, Nan Huang, Guanghui Hu, Xin Yang, Guoan Tang, LiYang Xiong, and Fayuan Li. 2019. “Integrated edge detection and terrain analysis for agricultural terrace delineation from remote sensing images.” *International Journal of Geographical Information Science* 34 (August): 1–20. <https://doi.org/10.1080/13658816.2019.1650363>.
- Davis, Larry S. 1975. “A survey of edge detection techniques” [in en]. *Computer Graphics and Image Processing* 4, no. 3 (September): 248–270. ISSN: 0146-664X, accessed March 3, 2023. [https://doi.org/10.1016/0146-664X\(75\)90012-X](https://doi.org/10.1016/0146-664X(75)90012-X). <https://www.sciencedirect.com/science/article/pii/0146664X7590012X>.
- Demirović, Damir. 2019. “An Implementation of the Mean Shift Algorithm” [in en]. *Image Processing On Line* 9 (September): 251–268. ISSN: 2105-1232, accessed March 8, 2023. <https://doi.org/10.5201/ipol.2019.255>. [https://www.ipol.im/pub/art/2019/255/?utm\\_source=doi](https://www.ipol.im/pub/art/2019/255/?utm_source=doi).
- Dey, V, Yun Zhang, and Ming Zhong. 2010. “A review on image segmentation techniques with remote sensing perspective.” *ISPRS TC VII Symposium - 100 Years ISPRS* 38 (January).
- Diakogiannis, Foivos I., François Waldner, Peter Caccetta, and Chen Wu. 2020. “ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data.” *ISPRS Journal of Photogrammetry and Remote Sensing* 162 (April): 94–114. ISSN: 0924-2716, accessed October 19, 2022. <https://doi.org/10.1016/j.isprsjprs.2020.01.013>. <http://www.scopus.com/inward/record.url?scp=85079865086&partnerID=8YFLogxK>.
- Ding, Lijun, and Ardesbir Goshtasby. 2001. “On the Canny edge detector” [in en]. *Pattern Recognition* 34, no. 3 (March): 721–725. ISSN: 0031-3203, accessed February 28, 2023. [https://doi.org/10.1016/S0031-3203\(00\)00023-6](https://doi.org/10.1016/S0031-3203(00)00023-6). <https://www.sciencedirect.com/science/article/pii/S0031320300000236>.
- Duro, Dennis C., Steven E. Franklin, and Monique G. Dubé. 2012. “A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the

- classification of agricultural landscapes using SPOT-5 HRG imagery” [in en]. *Remote Sensing of Environment* 118 (March): 259–272. ISSN: 0034-4257, accessed November 25, 2022. <https://doi.org/10.1016/j.rse.2011.11.020>. <https://www.sciencedirect.com/science/article/pii/S0034425711004172>.
- Eastwood, Robert, Michael Lipton, and Andrew Newell. 2010. “Chapter 65 Farm Size.” In *Handbook of Agricultural Economics*, 4:3323–3397. Elsevier, January. Accessed August 19, 2023. [https://doi.org/10.1016/S1574-0072\(09\)04065-1](https://doi.org/10.1016/S1574-0072(09)04065-1). <https://www.sciencedirect.com/science/article/pii/S1574007209040651>.
- Economist, The. 2017. *Global Food Security Index*. Accessed August 19, 2023. [https://www.eiu.com/public/topical\\_report.aspx?campaignid=GFSI2017](https://www.eiu.com/public/topical_report.aspx?campaignid=GFSI2017).
- Geospatial, Trimble. 2022. *What is eCognition? / Trimble Geospatial*. Accessed November 21, 2022. <https://geospatial.trimble.com/what-is-ecognition>.
- Glorot, Xavier, and Yoshua Bengio. 2010. “Understanding the difficulty of training deep feedforward neural networks” [in en]. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, 249–256. ISSN: 1938-7228. JMLR Workshop / Conference Proceedings, March. Accessed July 5, 2023. <https://proceedings.mlr.press/v9/glorot10a.html>.
- Guo, Yanming, Yu Liu, Theodoros Georgiou, and Michael S. Lew. 2018. “A review of semantic segmentation using deep neural networks” [in en]. *International Journal of Multimedia Information Retrieval* 7, no. 2 (June): 87–93. ISSN: 2192-662X, accessed October 18, 2022. <https://doi.org/10.1007/s13735-017-0141-z>. <https://doi.org/10.1007/s13735-017-0141-z>.
- Hao, Shijie, Yuan Zhou, and Yanrong Guo. 2020. “A Brief Survey on Semantic Segmentation with Deep Learning” [in en]. *Neurocomputing* 406 (September): 302–321. ISSN: 0925-2312, accessed October 18, 2022. <https://doi.org/10.1016/j.neucom.2019.11.118>. <https://www.sciencedirect.com/science/article/pii/S0925231220305476>.
- Hay, G. J., and G. Castilla. 2008. “Geographic Object-Based Image Analysis (GEOBIA): A new name for a new discipline” [in en]. In *Object-Based Image Analysis: Spatial Concepts for Knowledge-Driven Remote Sensing Applications*, edited by Thomas Blaschke, Stefan Lang, and Geoffrey J. Hay, 75–89. Lecture Notes in Geoinformation and Cartography. Berlin, Heidelberg: Springer. ISBN: 978-3-540-77058-9, accessed November 18, 2022. [https://doi.org/10.1007/978-3-540-77058-9\\_4](https://doi.org/10.1007/978-3-540-77058-9_4). [https://doi.org/10.1007/978-3-540-77058-9\\_4](https://doi.org/10.1007/978-3-540-77058-9_4).
- He, Kaiming, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. 2018. *Mask R-CNN*. ArXiv:1703.06870 [cs] version: 3, January. Accessed July 21, 2023. <http://arxiv.org/abs/1703.06870>.
- He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. “Identity Mappings in Deep Residual Networks” [in en]. In *Computer Vision – ECCV 2016*, edited by Bastian Leibe, Jiri

- Matas, Nicu Sebe, and Max Welling, 630–645. Lecture Notes in Computer Science. Cham: Springer International Publishing. ISBN: 978-3-319-46493-0. [https://doi.org/10.1007/978-3-319-46493-0\\_38](https://doi.org/10.1007/978-3-319-46493-0_38).
- Heath, M.D., S. Sarkar, T. Sanocki, and K.W. Bowyer. 1997. “A robust visual method for assessing the relative performance of edge-detection algorithms.” Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19, no. 12 (December): 1338–1359. ISSN: 1939-3539. <https://doi.org/10.1109/34.643893>.
- Hecht-nielsen, ROBERT. 1992. “III.3 - Theory of the Backpropagation Neural Network\*\*Based on “nonindent” by Robert Hecht-Nielsen, which appeared in Proceedings of the International Joint Conference on Neural Networks 1, 593–611, June 1989. © 1989 IEEE.” [in en]. In *Neural Networks for Perception*, edited by Harry Wechsler, 65–93. Academic Press, January. ISBN: 978-0-12-741252-8, accessed July 21, 2023. <https://doi.org/10.1016/B978-0-12-741252-8.50010-8>. <https://www.sciencedirect.com/science/article/pii/B9780127412528500108>.
- INSTAT. 2023. *Institut National de la Statistique du Mali/INSTAT* [in fr]. Accessed July 2, 2023. <https://instat-mali.org/fr>.
- Jozdani, Shahab, and Dongmei Chen. 2020. “On the versatility of popular and recently proposed supervised evaluation metrics for segmentation quality of remotely sensed images: An experimental case study of building extraction” [in en]. *ISPRS Journal of Photogrammetry and Remote Sensing* 160 (February): 275–290. ISSN: 0924-2716, accessed May 16, 2023. <https://doi.org/10.1016/j.isprsjprs.2020.01.002>. <https://www.sciencedirect.com/science/article/pii/S0924271620300022>.
- Khatami, Reza, Giorgos Mountrakis, and Stephen V. Stehman. 2016. “A meta-analysis of remote sensing research on supervised pixel-based land-cover image classification processes: General guidelines for practitioners and future research” [in en]. *Remote Sensing of Environment* 177 (May): 89–100. ISSN: 0034-4257, accessed November 22, 2022. <https://doi.org/10.1016/j.rse.2016.02.028>. <https://www.sciencedirect.com/science/article/pii/S0034425716300578>.
- Kingma, Diederik P., and Jimmy Ba. 2017. *Adam: A Method for Stochastic Optimization*. ArXiv:1412.6980 [cs], January. Accessed July 5, 2023. <https://doi.org/10.48550/arXiv.1412.6980>. <http://arxiv.org/abs/1412.6980>.
- Kotaridis, Ioannis, and Maria Lazaridou. 2021. “Remote sensing image segmentation advances: A meta-analysis” [in en]. *ISPRS Journal of Photogrammetry and Remote Sensing* 173 (March): 309–322. ISSN: 0924-2716, accessed October 19, 2022. <https://doi.org/10.1016/j.isprsjprs.2021.03.002>.



isprsjprs.2021.01.020. <https://www.sciencedirect.com/science/article/pii/S0924271621000265>.

Krähenbühl, Philipp, and Vladlen Koltun. 2011. “Efficient Inference in Fully Connected CRFs with Gaussian Edge Potentials.” In *Advances in Neural Information Processing Systems*, vol. 24. Curran Associates, Inc. Accessed May 13, 2023. [https://proceedings.neurips.cc/paper\\_files/paper/2011/hash/beda24c1e1b46055dff2c39c98fd6fc1-Abstract.html](https://proceedings.neurips.cc/paper_files/paper/2011/hash/beda24c1e1b46055dff2c39c98fd6fc1-Abstract.html).

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. 2017. “ImageNet classification with deep convolutional neural networks.” *Communications of the ACM* 60, no. 6 (May): 84–90. ISSN: 0001-0782, accessed May 8, 2023. <https://doi.org/10.1145/3065386>. <https://dl.acm.org/doi/10.1145/3065386>.

Lambert, Marie-Julie. 2018. “Estimating and understanding smallholder crop production variability from pixel to village levels thanks to satellite observation in Mali” [in en]. PhD diss., UCL - Université Catholique de Louvain. Accessed October 3, 2022. <https://dial.uclouvain.be/pr/boreal/object/boreal:207897>.

LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. 2015. “Deep learning” [in en]. Number: 7553 Publisher: Nature Publishing Group, *Nature* 521, no. 7553 (May): 436–444. ISSN: 1476-4687, accessed May 4, 2023. <https://doi.org/10.1038/nature14539>. <https://www.nature.com/articles/nature14539>.

Li, Ke, Gang Wan, Gong Cheng, Liqiu Meng, and Junwei Han. 2020. “Object detection in optical remote sensing images: A survey and a new benchmark” [in en]. *ISPRS Journal of Photogrammetry and Remote Sensing* 159 (January): 296–307. ISSN: 09242716, accessed November 4, 2022. <https://doi.org/10.1016/j.isprsjprs.2019.11.023>. <https://linkinghub.elsevier.com/retrieve/pii/S0924271619302825>.

Long, Jonathan, Evan Shelhamer, and Trevor Darrell. 2015. “Fully Convolutional Networks for Semantic Segmentation,” 3431–3440. Accessed May 8, 2023. [https://openaccess.thecvf.com/content\\_cvpr\\_2015/html/Long\\_Fully\\_Convolutional\\_Networks\\_2015\\_CVPR\\_paper.html](https://openaccess.thecvf.com/content_cvpr_2015/html/Long_Fully_Convolutional_Networks_2015_CVPR_paper.html).

Lowder, Sarah K., Jakob Skoet, and Terri Raney. 2016. “The Number, Size, and Distribution of Farms, Smallholder Farms, and Family Farms Worldwide.” *World Development* 87 (November): 16–29. ISSN: 0305-750X, accessed August 19, 2023. <https://doi.org/10.1016/j.worlddev.2015.10.041>. <https://www.sciencedirect.com/science/article/pii/S0305750X15002703>.

Lu, D., and Q. Weng. 2007. “A survey of image classification methods and techniques for improving classification performance.” Publisher: Taylor & Francis \_eprint: <https://doi.org/10.1080/01431160> *International Journal of Remote Sensing* 28, no. 5 (March): 823–870. ISSN: 0143-1161, ac-

- cessed November 23, 2022. <https://doi.org/10.1080/01431160600746456>. <https://doi.org/10.1080/01431160600746456>.
- Lucieer, A., and A. Stein. 2002. “Existential uncertainty of spatial objects segmented from satellite sensor imagery.” Conference Name: IEEE Transactions on Geoscience and Remote Sensing, *IEEE Transactions on Geoscience and Remote Sensing* 40, no. 11 (November): 2518–2521. ISSN: 1558-0644. <https://doi.org/10.1109/TGRS.2002.805072>.
- Ma, Lei, Yu Liu, Xueliang Zhang, Yuanxin Ye, Gaofer Yin, and Brian Alan Johnson. 2019. “Deep learning in remote sensing applications: A meta-analysis and review” [in en]. *ISPRS Journal of Photogrammetry and Remote Sensing* 152 (June): 166–177. ISSN: 0924-2716, accessed October 3, 2022. <https://doi.org/10.1016/j.isprsjprs.2019.04.015>. <https://www.sciencedirect.com/science/article/pii/S0924271619301108>.
- Mali: Institut National de la Statistique, République de. 2012. *Resultats RGPH 2009 (Région de Sikasso)* [in French], July. Accessed June 30, 2023. <https://web.archive.org/web/20120727120405/http://instat.gov.ml/documentation/sikasso.pdf>.
- Maxar. 2023. *Earth Intelligence & Space Infrastructure*. Accessed July 20, 2023. <https://www.maxar.com/>.
- Meyer, Fernand. 2012. *The watershed concept and its use in segmentation : a brief history*. ArXiv:1202.0216 [cs], February. Accessed May 13, 2023. <https://doi.org/10.48550/arXiv.1202.0216>. <http://arxiv.org/abs/1202.0216>.
- Mo, Yujian, Yan Wu, Xinneng Yang, Feilin Liu, and Yujun Liao. 2022. “Review the state-of-the-art technologies of semantic segmentation based on deep learning” [in en]. *Neurocomputing* 493 (July): 626–646. ISSN: 0925-2312, accessed May 14, 2023. <https://doi.org/10.1016/j.neucom.2022.01.005>. <https://www.sciencedirect.com/science/article/pii/S0925231222000054>.
- Neupane, Bipul, Teerayut Horanont, and Jagannath Aryal. 2021. “Deep Learning-Based Semantic Segmentation of Urban Features in Satellite Images: A Review and Meta-Analysis” [in en]. Number: 4 Publisher: Multidisciplinary Digital Publishing Institute, *Remote Sensing* 13, no. 4 (January): 808. ISSN: 2072-4292, accessed October 3, 2022. <https://doi.org/10.3390/rs13040808>. <https://www.mdpi.com/2072-4292/13/4/808>.
- Nguyen, T. B, and D Ziou. 2000. “Contextual and non-contextual performance evaluation of edge detectors” [in en]. *Pattern Recognition Letters* 21, no. 9 (August): 805–816. ISSN: 0167-8655, accessed March 3, 2023. [https://doi.org/10.1016/S0167-8655\(00\)00045-3](https://doi.org/10.1016/S0167-8655(00)00045-3). <https://www.sciencedirect.com/science/article/pii/S0167865500000453>.
- Opitz, D., and S. Blundell. 2008. “Object recognition and image segmentation: the Feature Analyst® approach” [in en]. In *Object-Based Image Analysis: Spatial Concepts for Knowledge-Driven Remote Sensing Applications*, edited by Thomas Blaschke, Stefan Lang, and Geof-

- frey J. Hay, 153–167. *Lecture Notes in Geoinformation and Cartography*. Berlin, Heidelberg: Springer. ISBN: 978-3-540-77058-9, accessed November 20, 2022. [https://doi.org/10.1007/978-3-540-77058-9\\_8](https://doi.org/10.1007/978-3-540-77058-9_8). [https://doi.org/10.1007/978-3-540-77058-9\\_8](https://doi.org/10.1007/978-3-540-77058-9_8).
- Pal, Nikhil R, and Sankar K Pal. 1993. “A review on image segmentation techniques” [in en]. *Pattern Recognition* 26, no. 9 (September): 1277–1294. ISSN: 0031-3203, accessed October 19, 2022. [https://doi.org/10.1016/0031-3203\(93\)90135-J](https://doi.org/10.1016/0031-3203(93)90135-J). <https://www.sciencedirect.com/science/article/pii/003132039390135J>.
- Pavlova, Maria, Valerii Timofeev, Dmitry Bocharov, Irina Kunina, Anna Smagina, and Mikhail Zagarev. 2021. *Segmentation of Agricultural Parcels in Satellite Images Based on Historical Vegetation Index Data*. December. <https://doi.org/10.20944/preprints202112.0243.v1>.
- Pellegrino, F.A., W. Vanzella, and V. Torre. 2004. “Edge detection revisited.” Conference Name: IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 34, no. 3 (June): 1500–1518. ISSN: 1941-0492. <https://doi.org/10.1109/TSMCB.2004.824147>.
- Persello, C., V. A. Tolpekin, J. R. Bergado, and R. A. de By. 2019. “Delineation of agricultural fields in smallholder farms from satellite images using fully convolutional networks and combinatorial grouping” [in en]. *Remote Sensing of Environment* 231 (September): 111253. ISSN: 0034-4257, accessed October 3, 2022. <https://doi.org/10.1016/j.rse.2019.111253>. <https://www.sciencedirect.com/science/article/pii/S003442571930272X>.
- Pham, Dzung, Chenyang Xu, and Jerry Prince. 2000. “A Survey of Current Methods in Medical Image Segmentation.” *Annual review of biomedical engineering* 2 (February): 315–37. <https://doi.org/10.1146/annurev.bioeng.2.1.315>.
- Planet. 2023. *Planet Basemaps: Comprehensive, High-Frequency Mosaics for Analysis* [in en]. Accessed July 18, 2023. <https://www.planet.com/products/basemap/>.
- Qi, Wenwen, Mengfei Wei, Wentao Yang, Chong Xu, and Chao Ma. 2020. “Automatic Mapping of Landslides by the ResU-Net” [in en]. Number: 15 Publisher: Multidisciplinary Digital Publishing Institute, *Remote Sensing* 12, no. 15 (January): 2487. ISSN: 2072-4292, accessed October 12, 2022. <https://doi.org/10.3390/rs12152487>. <https://www.mdpi.com/2072-4292/12/15/2487>.
- Rong, Weibin, Zhanjing Li, Wei Zhang, and Lining Sun. 2014. “An improved Canny edge detection algorithm.” In *2014 IEEE International Conference on Mechatronics and Automation*, 577–582. ISSN: 2152-744X. August. <https://doi.org/10.1109/ICMA.2014.6885761>.
- Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. 2015. “U-Net: Convolutional Networks for Biomedical Image Segmentation” [in en]. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, edited by Nassir Navab, Joachim Hornegger, William

- M. Wells, and Alejandro F. Frangi, 234–241. Lecture Notes in Computer Science. Cham: Springer International Publishing. ISBN: 978-3-319-24574-4. [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28).
- Simonyan, Karen, and Andrew Zisserman. 2015. *Very Deep Convolutional Networks for Large-Scale Image Recognition*. ArXiv:1409.1556 [cs], April. Accessed May 8, 2023. <https://doi.org/10.48550/arXiv.1409.1556>. <http://arxiv.org/abs/1409.1556>.
- Song, Hwanjun, Minseok Kim, Dongmin Park, Yooju Shin, and Jae-Gil Lee. 2022. “Learning From Noisy Labels With Deep Neural Networks: A Survey.” Conference Name: IEEE Transactions on Neural Networks and Learning Systems, *IEEE Transactions on Neural Networks and Learning Systems*, 1–19. ISSN: 2162-2388. <https://doi.org/10.1109/TNNLS.2022.3152527>.
- Spirkovska, Lilly. 1993. *A Summary of Image Segmentation*. Techniques Ames Research Center. [https://books.google.at/books?id=BLA3AQAAMAAJ&dq=spirkovska+1993&lr=&hl=fr&source=gbs\\_navlinks\\_s](https://books.google.at/books?id=BLA3AQAAMAAJ&dq=spirkovska+1993&lr=&hl=fr&source=gbs_navlinks_s).
- Sultana, Farhana, Abu Sufian, and Paramartha Dutta. 2020. “Evolution of Image Segmentation using Deep Convolutional Neural Network: A Survey.” ArXiv:2001.04074 [cs], *Knowledge-Based Systems* 201-202 (August): 106062. ISSN: 09507051, accessed October 21, 2022. <https://doi.org/10.1016/j.knosys.2020.106062>. <http://arxiv.org/abs/2001.04074>.
- Szegedy, Christian, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. 2015. “Going Deeper With Convolutions,” 1–9. Accessed May 19, 2023. [https://www.cv-foundation.org/openaccess/content\\_cvpr\\_2015/html/Szegedy\\_Going\\_Deeper\\_With\\_2015\\_CVPR\\_paper.html](https://www.cv-foundation.org/openaccess/content_cvpr_2015/html/Szegedy_Going_Deeper_With_2015_CVPR_paper.html).
- USDA. 2023. *USDA Foreign Agricultural Service* [in en], June. Accessed June 27, 2023. <https://fas.usda.gov/home>.
- USGS. 2023. *What are the best Landsat spectral bands for use in my research? | U.S. Geological Survey*. Accessed August 20, 2023. <https://www.usgs.gov/faqs/what-are-best-landsat-spectral-bands-use-my-research>.
- Waldner, François, and Foivos I. Diakogiannis. 2020. *Deep learning on edge: extracting field boundaries from satellite images with a convolutional neural network*. ArXiv:1910.12023 [cs], February. Accessed November 8, 2022. <https://doi.org/10.48550/arXiv.1910.12023>. <http://arxiv.org/abs/1910.12023>.
- Wan, Xing. 2019. “Influence of feature scaling on convergence of gradient iterative algorithm” [in en]. Publisher: IOP Publishing, *Journal of Physics: Conference Series* 1213, no. 3 (June): 032021. ISSN: 1742-6596, accessed July 19, 2023. <https://doi.org/10.1088/1742-6596/1213/3/032021>. <https://dx.doi.org/10.1088/1742-6596/1213/3/032021>.

- Wang, Bulent Bayram, and Elif Sertel. 2022. “A comprehensive review on deep learning based remote sensing image super-resolution methods” [in en]. *Earth-Science Reviews* 232 (September): 104110. ISSN: 0012-8252, accessed October 3, 2022. <https://doi.org/10.1016/j.earscir.2022.104110>. <https://www.sciencedirect.com/science/article/pii/S0012825222001945>.
- Wang, E. Wang, and Ying Zhu. 2020. “Image segmentation evaluation: a survey of methods” [in en]. *Artificial Intelligence Review* 53, no. 8 (December): 5637–5674. ISSN: 1573-7462, accessed May 14, 2023. <https://doi.org/10.1007/s10462-020-09830-9>. <https://doi.org/10.1007/s10462-020-09830-9>.
- Weidner, Uwe. 2008. “Contribution to the assessment of segmentation quality for remote sensing applications.” *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 37 (B7): 479–484.
- Yan, L., and D.P. Roy. 2014. “Automated crop field extraction from multi-temporal Web Enabled Landsat Data” [in en]. *Remote Sensing of Environment* 144 (March): 42–64. ISSN: 00344257, accessed July 21, 2023. <https://doi.org/10.1016/j.rse.2014.01.006>. <https://linkinghub.elsevier.com/retrieve/pii/S0034425714000194>.
- Yang, Jian, Yuhong He, John Caspersen, and Trevor Jones. 2015. “A discrepancy measure for segmentation evaluation from the perspective of object recognition” [in en]. *ISPRS Journal of Photogrammetry and Remote Sensing* 101 (March): 186–192. ISSN: 0924-2716, accessed May 20, 2023. <https://doi.org/10.1016/j.isprsjprs.2014.12.015>. <https://www.sciencedirect.com/science/article/pii/S0924271614002901>.
- Yuan, Xiaohui, Jianfang Shi, and Lichuan Gu. 2021. “A review of deep learning methods for semantic segmentation of remote sensing imagery” [in en]. *Expert Systems with Applications* 169 (May): 114417. ISSN: 0957-4174, accessed October 12, 2022. <https://doi.org/10.1016/j.eswa.2020.114417>. <https://www.sciencedirect.com/science/article/pii/S0957417420310836>.
- Zhang, Xueliang, Xuezhi Feng, Pengfeng Xiao, Guangjun He, and Liujun Zhu. 2015. “Segmentation quality evaluation using region-based precision and recall measures for remote sensing images” [in en]. *ISPRS Journal of Photogrammetry and Remote Sensing* 102 (April): 73–84. ISSN: 0924-2716, accessed May 19, 2023. <https://doi.org/10.1016/j.isprsjprs.2015.01.009>. <https://www.sciencedirect.com/science/article/pii/S092427161500026X>.
- Zhang, Y. J. 1996. “A survey on evaluation methods for image segmentation” [in en]. *Pattern Recognition* 29, no. 8 (August): 1335–1346. ISSN: 0031-3203, accessed October 21, 2022. [https://doi.org/10.1016/0031-3203\(95\)00169-7](https://doi.org/10.1016/0031-3203(95)00169-7). <https://www.sciencedirect.com/science/article/pii/0031320395001697>.
- Zhao, Hengshuang, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. 2017. “Pyramid Scene Parsing Network,” 2881–2890. Accessed June 29, 2023. <https://openaccess.thecvf>.



[com/content\\_cvpr\\_2017/html/Zhao\\_Pyramid\\_Scene\\_Parsing\\_CVPR\\_2017\\_paper.html](https://arxiv.org/abs/1703.07504).

Zhenrong, Du, Yang Jianyu, Ou Cong, and Zhang Tingting. 2019. *Smallholder Crop Area Mapped with a Semantic Segmentation Deep Learning Method*, April. Accessed July 22, 2023. <https://www.mdpi.com/2072-4292/11/7/888>.

Zhong, Bo, Tengfei Wei, Xiaobo Luo, Bailin Du, Longfei Hu, Kai Ao, Aixia Yang, and Junjun Wu. 2023. “Multi-Swin Mask Transformer for Instance Segmentation of Agricultural Field Extraction” [in en]. Number: 3 Publisher: Multidisciplinary Digital Publishing Institute, *Remote Sensing* 15, no. 3 (January): 549. ISSN: 2072-4292, accessed July 21, 2023. <https://doi.org/10.3390/rs15030549>. <https://www.mdpi.com/2072-4292/15/3/549>.

## Appendix A: Additional material and methods figures

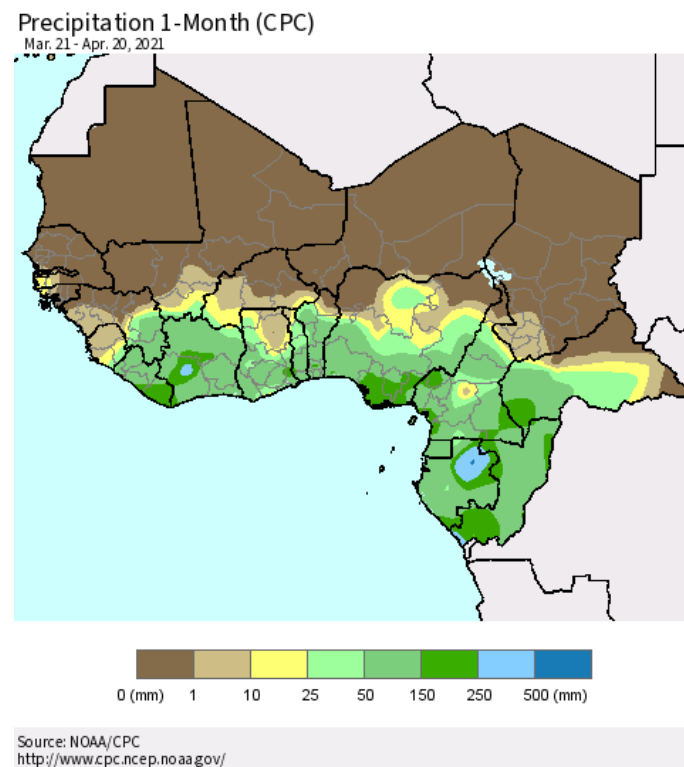


Figure A.1: 1-month Precipitation according to CPC (Mar. 21 - Apr. 20, 2021).

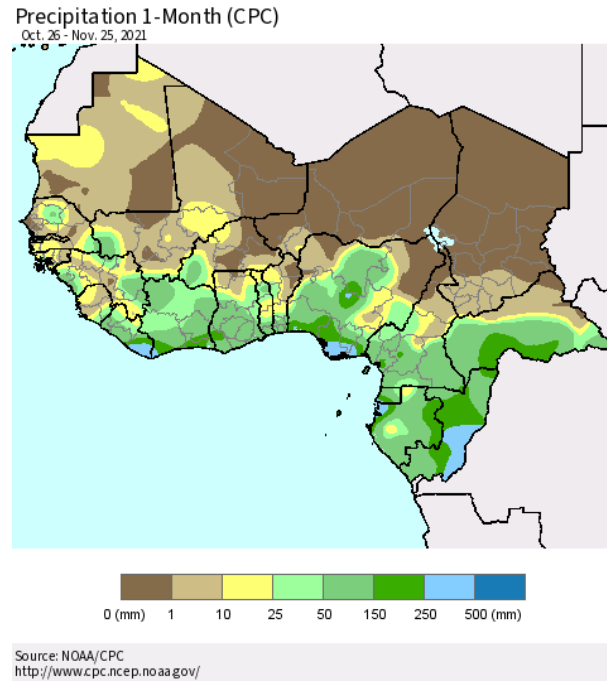


Figure A.2: 1-month Precipitation according to CPC (Oct. 26 - Nov. 25, 2021).

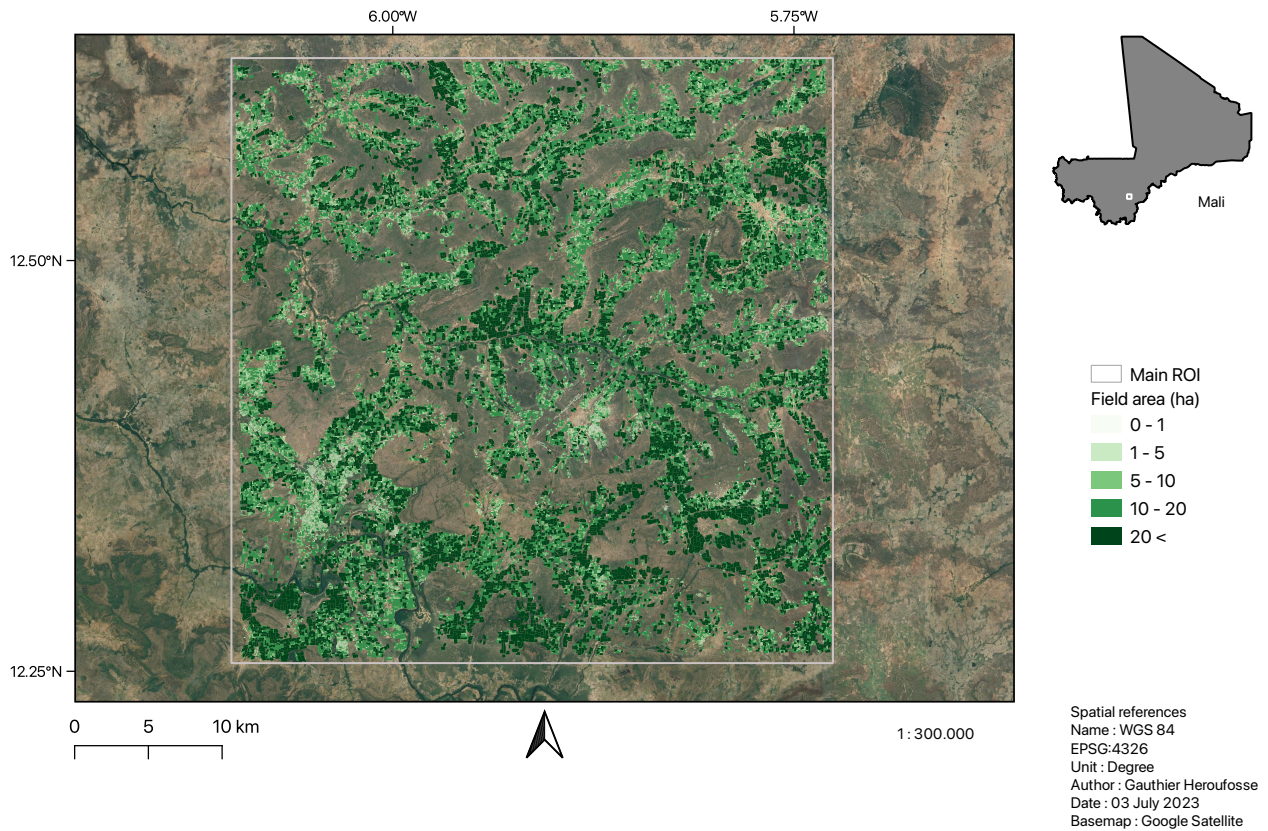


Figure A.3: Complete quality controlled reference dataset displaying the field size of each individual polygon.

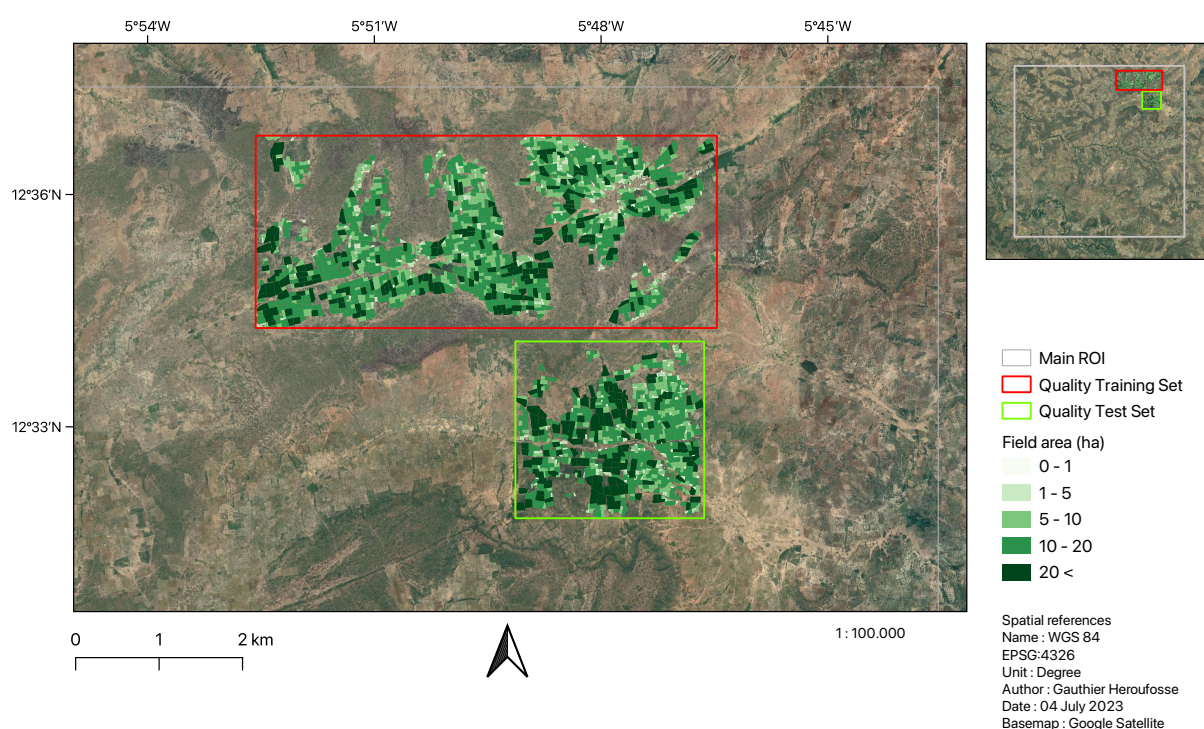


Figure A.4: Complete Quality set displaying the field size of each individual polygon.

## Appendix B: Results of the baseline method

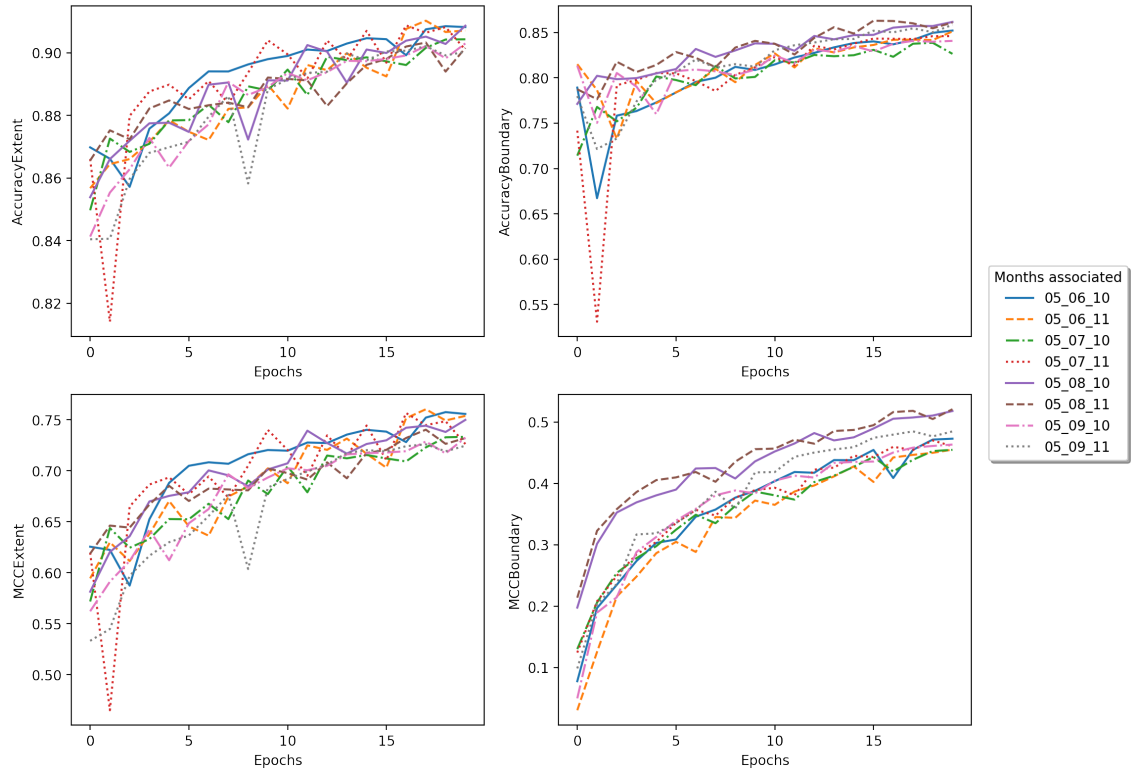


Figure B.1: Validation metrics comparison obtained with different monthly associations.



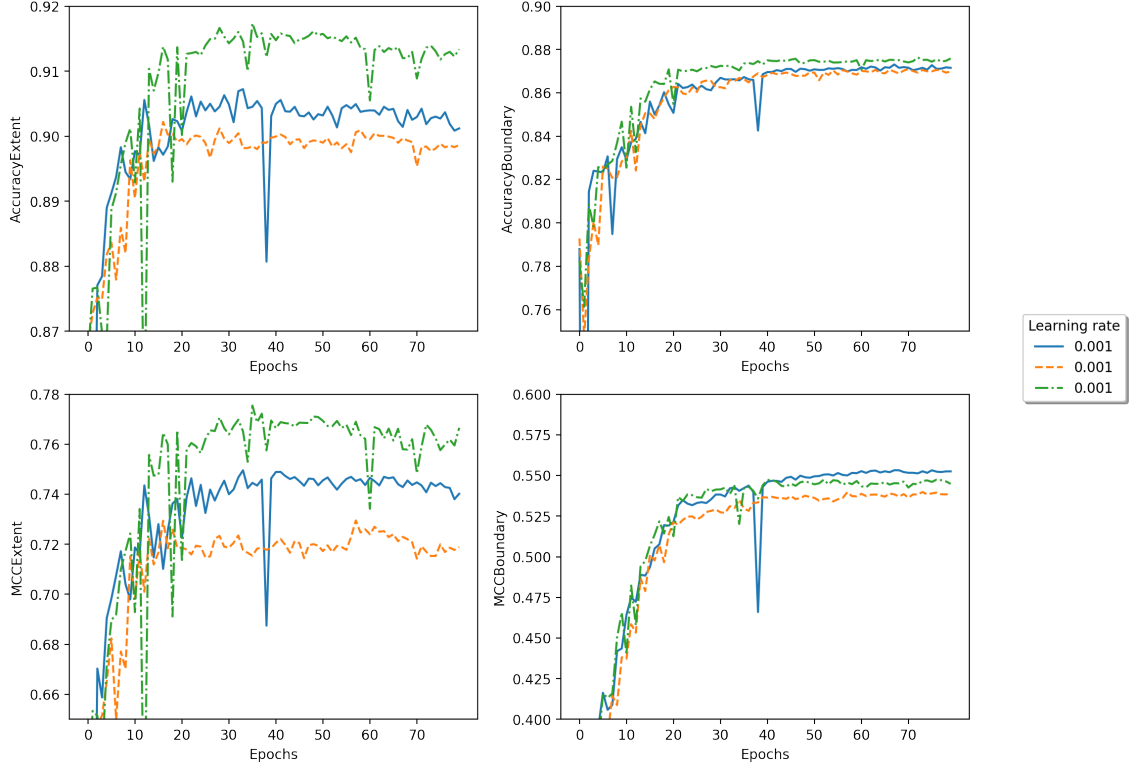


Figure B.2: Analysis of the variability of the validation metrics using the same LR for every model.

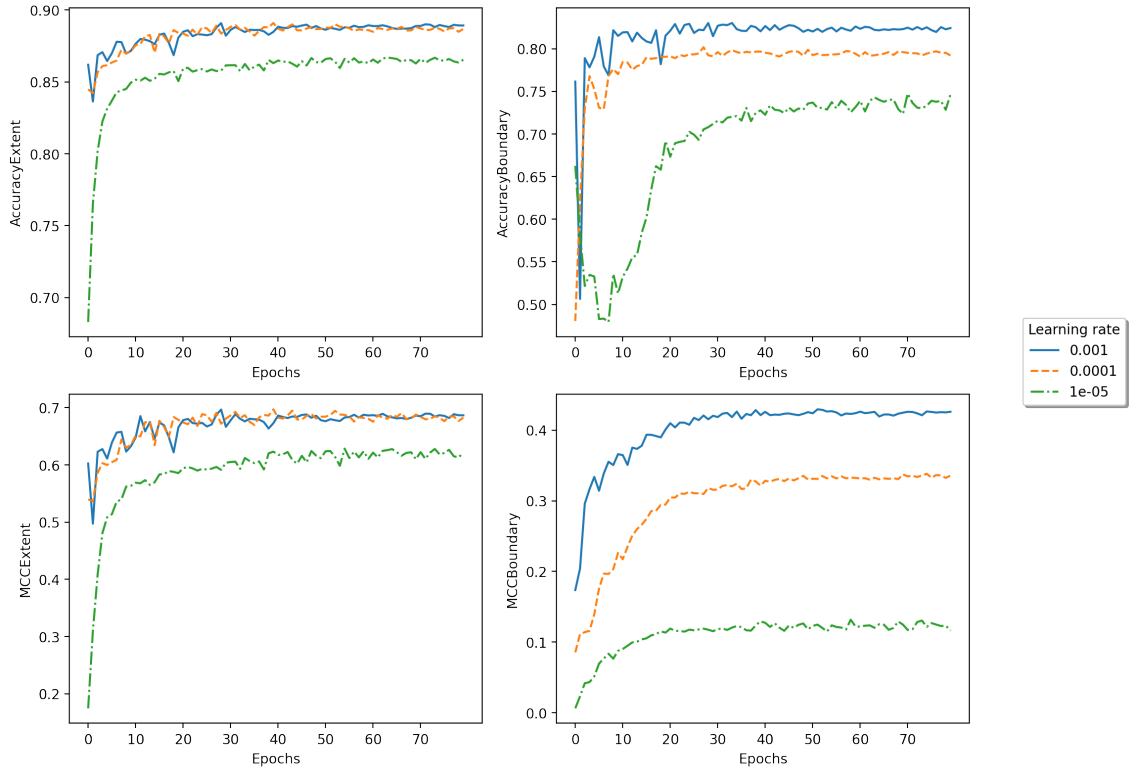


Figure B.3: Influence of the starting LR on the validation metrics.

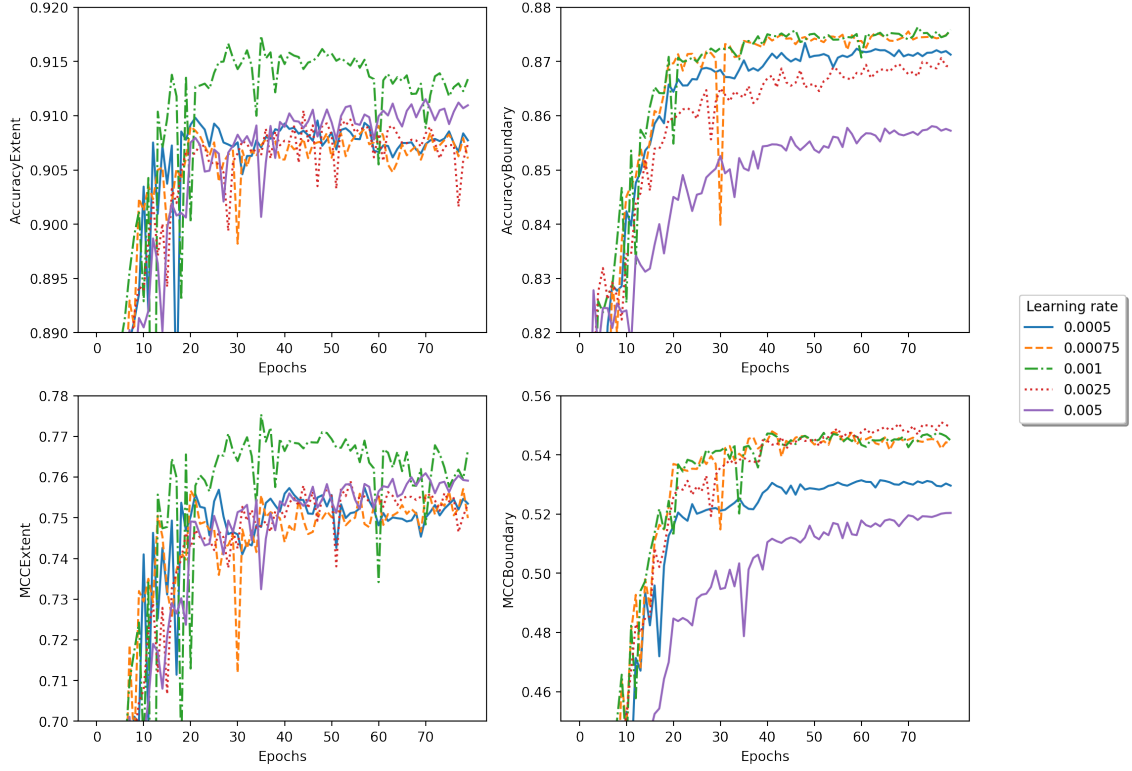


Figure B.4: Influence of the starting LR on the validation metrics, zoom around a LR of  $10^{-3}$ .

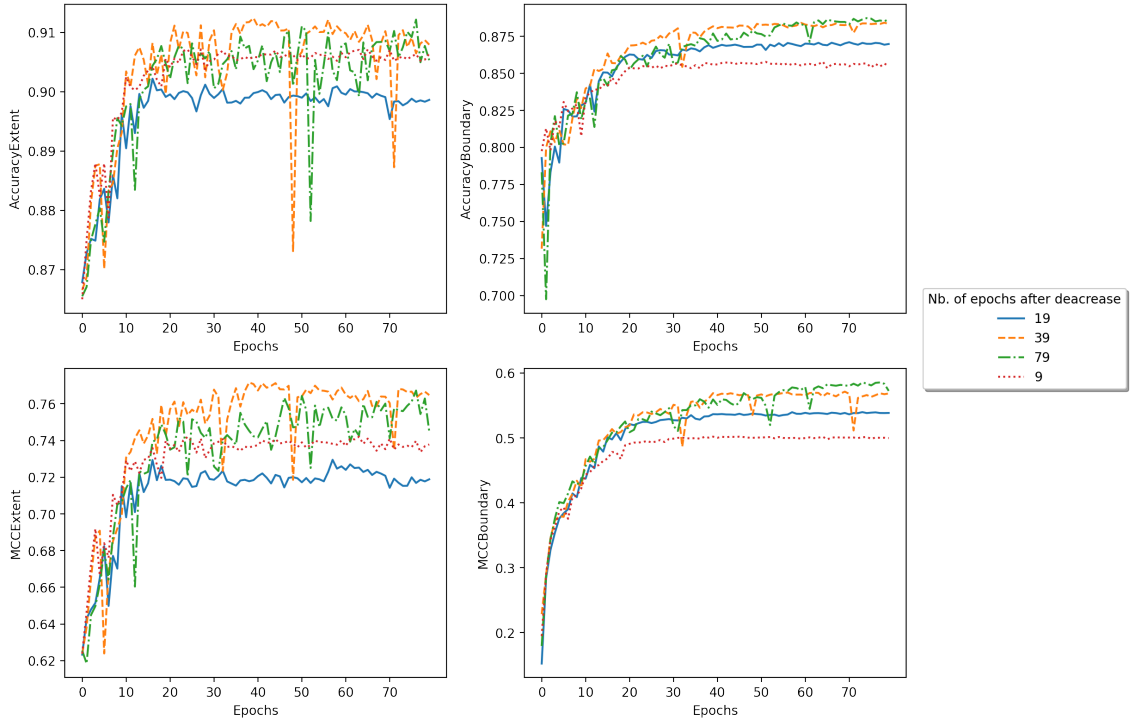


Figure B.5: Scheduler parameters for the weight decays (0.31%). Number of epochs after which the LR decrease.

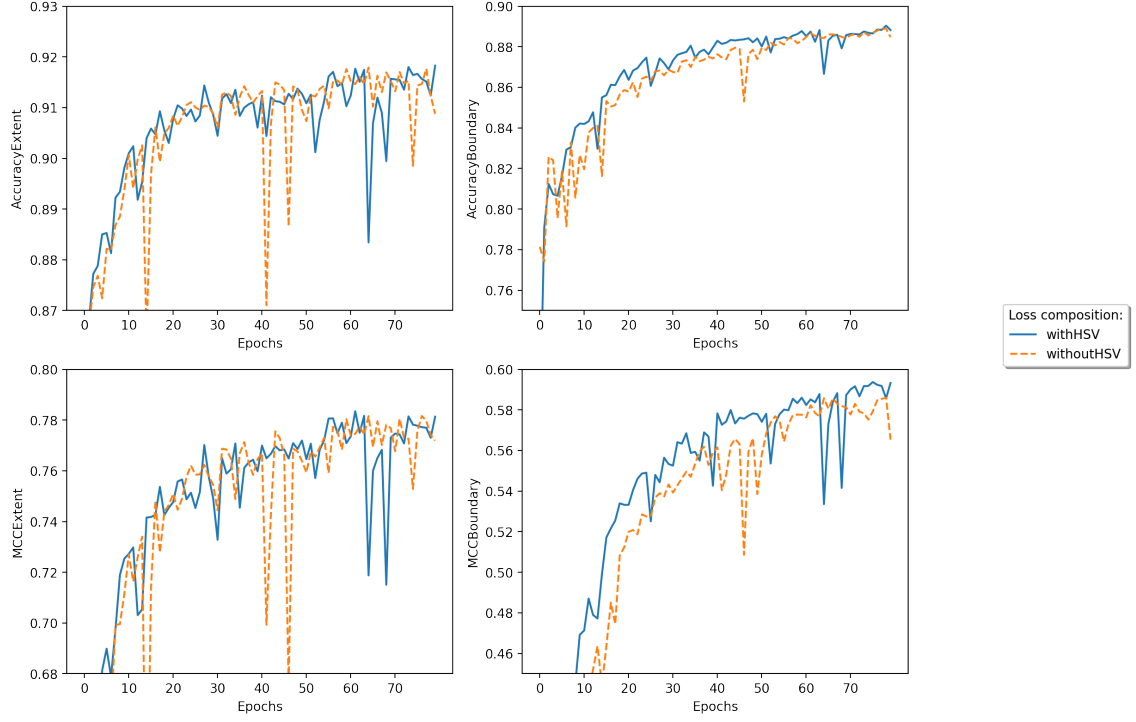


Figure B.6: Comparison of validation metrics with and without the HSV layer.

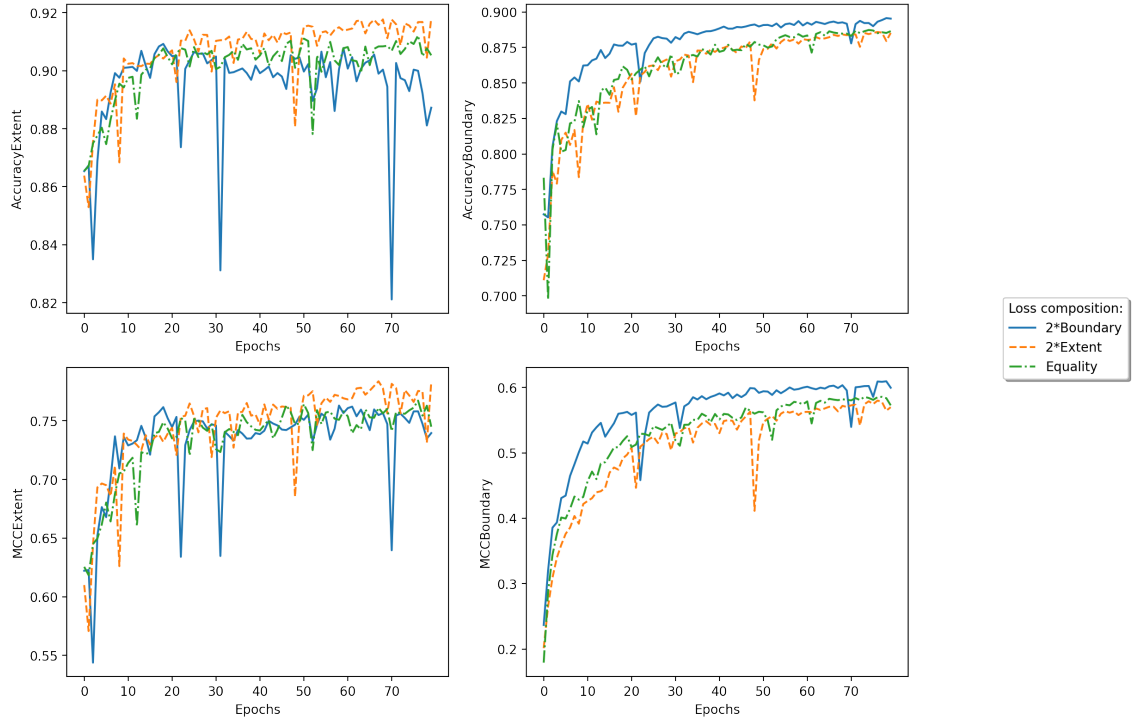


Figure B.7: Comparison of validation metrics obtained with enhanced recognition of boundaries and extent.

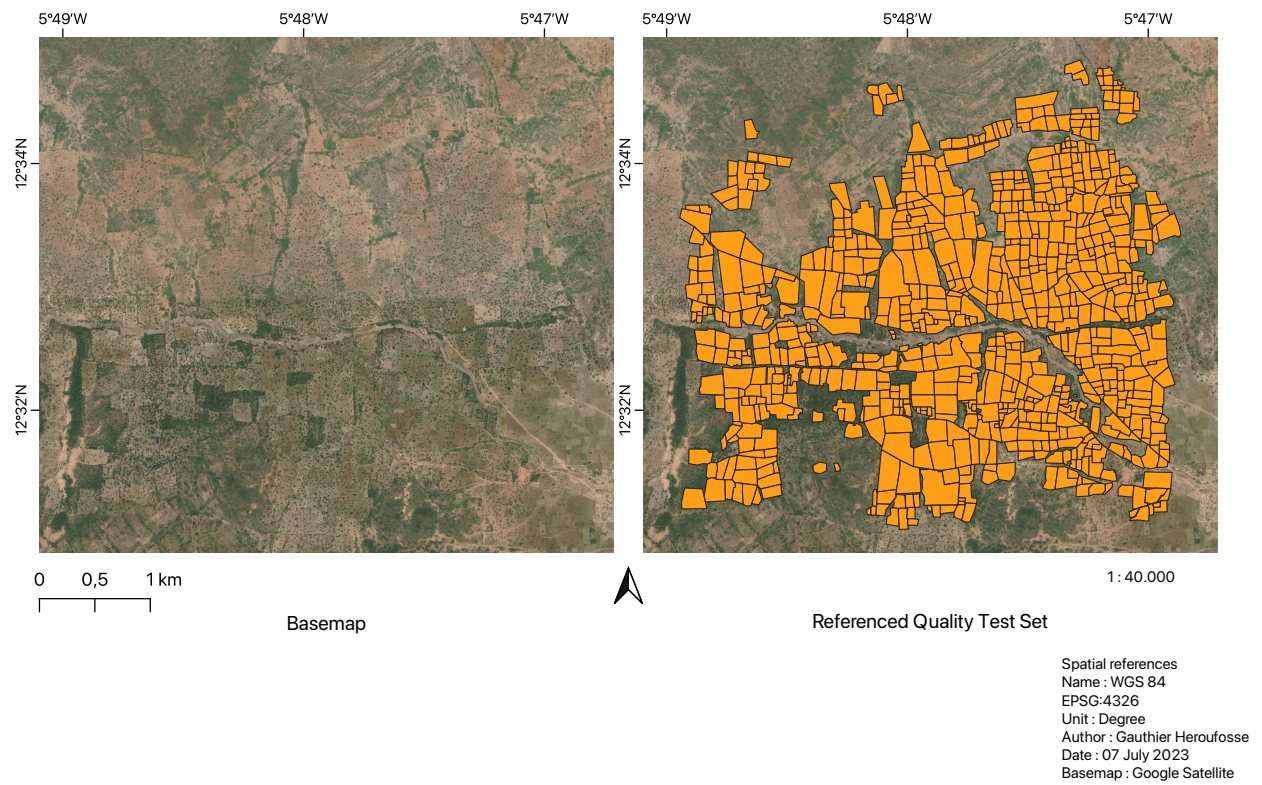


Figure B.8: Reference Quality Test Set

## Automatic field extraction using deep learning in the context of smallholder agriculture

Gauthier Heroufousse

Sub-Saharan Africa is the center of the world's population growth but is particularly vulnerable to food shortages. In the framework of collaboration between the Institut National de la Statistique du Mali (INSTAT) and UCLouvain, the goal is to achieve a better understanding of agricultural dynamics and improve agricultural statistics. Automatic field extraction from satellite imagery is an essential tool to achieve agricultural landscape mapping, which acts as a decision-support tool to help improve food security. A newly acquired set of reference agricultural field delimitations of 16.000km<sup>2</sup> has been prepared by the Malian partner of the Institut d'Economie Rurale (IER) in order to make the first step in that direction. This thesis aimed to evaluate the performance of an automatic field extraction methodology from High Resolution satellite images, in the context of smallholder agriculture, which is characteristic of the sub-Saharan region.

The methodology relies on a Deep Learning model trained with satellite images from 2021 captured by the Planet satellite and the reference dataset. The first step consisted of realizing a calibration analysis to set the best hyperparameters for the model training. The baseline method was then evaluated with carefully selected metrics showing the performance of our model. The last step includes an assessment of the model's portability within Africa, as well as an exploration of possible improvements by adjusting the quality of the input data and the imagery's spatial resolution.

Overall, the results were mixed. The model is mostly capable of finding the fields' extent but struggles to capture field boundaries correctly. The portability of the model is promising, but it suffers from land cover and topography differences. A major drawback is the lack of reliability of the reference dataset. Finally, recommendations and improvement opportunities were discussed for a potential continuation of experiments.