

# **A Remotely Sensed Based Comparison of Three Area Stratification Methods to Improve Estimation of Crop Area Statistics, A Case Study in Fragmented Landscapes of Ethiopia**

SHAFI DELSEBO BAMUD

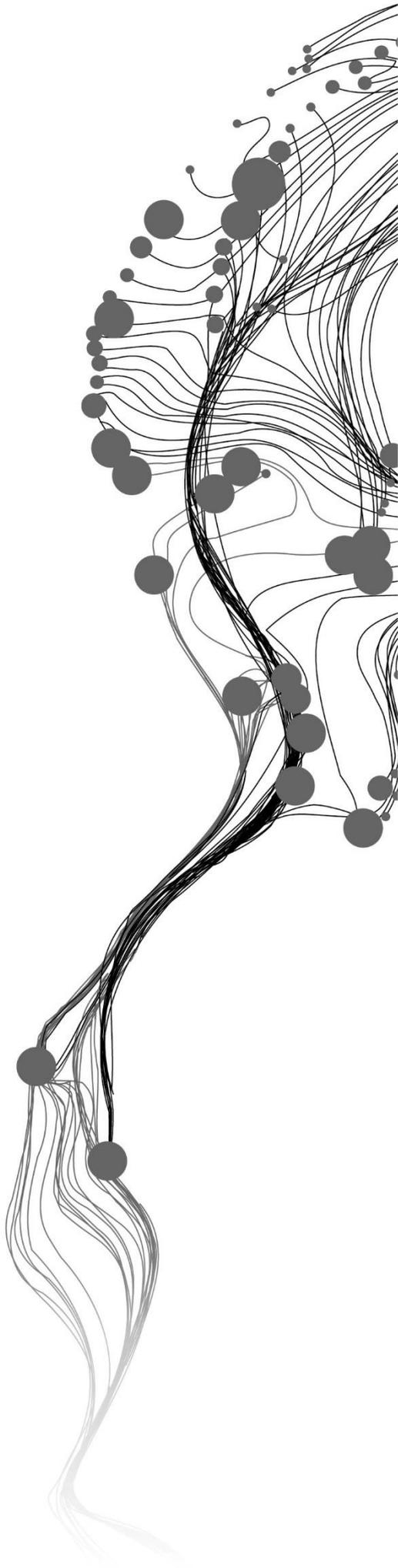
August,2022

SUPERVISORS:

Dr.ir. C.A.J.M. de Bie

Dr. C. Paris





# **A Remotely Sensed Based Comparison of Three Area Stratification Methods to Improve Estimation of Crop Area Statistics, A Case Study in Fragmented Landscapes of Ethiopia**

SHAFI DELSEBO BAMUD

Enschede, The Netherlands, August, 2022

Thesis submitted to the Faculty of Geo-Information Science and Earth  
Observation of the University of Twente in partial fulfilment of the  
requirements for the degree of Master of Science in Geo-information Science  
and Earth Observation.

Specialization: Specialization: Geoinformatics

## **SUPERVISORS:**

Dr.ir. C.A.J.M. de Bie

Dr. C. Paris

## **THESIS ASSESSMENT BOARD:**

Dr. ir. A. Vrieling (Chair)

Dr. B.H.P. Maathuis (External Examiner, University of Twente)

#### DISCLAIMER

This document describes work undertaken as part of a programme of study at the Faculty of Geo-Information Science and Earth Observation of the University of Twente. All views and opinions expressed therein remain the sole responsibility of the author, and do not necessarily represent those of the Faculty.

## ABSTRACT

Agricultural statistics, particularly crop production data, are crucial for food security analysis. Crop production is a product of crop area and yield. Thus, accurate and reliable crop area estimation is a critical component of agricultural statistics because it is vital in assessing food production. The optimal method of crop area is determined by various operational criteria, including land arrangement, crop field shape, crop type, cropping pattern, and available skills and resources. In Sub-Saharan African countries, farmlands are small, dispersed, and complex, making estimation of crop area challenging. In addition, high landscape fragmentation, the extreme variability of the environment over space and time and a mixed cropping system are challenging for crop area estimation. Many developing countries, especially sub-Saharan African countries, are affected by the early warning system of food security because of inadequate crop area statistics. Even though there have been various types of crop area estimates, the uncertainty of each type of crop area estimate varies with the topographic nature of the area and other related factors. Therefore, area stratification might be based on areas that do not properly represent the areas of actual crop fields, which leads to the biased representation of area stratification for crop area statistics in fragmented landscapes. To get an accurate and reliable crop area estimate, identifying the most appropriate area stratification method is very important, which in turn reduces the problem of food security that is caused by a lack of accurate crop area estimation. This research compared three existing area stratification methods for Oromia, Ethiopia to improve crop area estimation using high-resolution NDVI imagery. The three existing strata maps (CPS zone, LH zone and administrative unit map) were intersected to prepare a sampling scheme of Planet satellite imagery that specified clusters of tiles (1 km<sup>2</sup>) of sampled images. As many as 400 sample areas (tiles) were selected as sampling areas. The Planet monthly composite imagery of 2021 was used to obtain the time series (12 month) NDVI images of the 400 tiles. Then ISODATA unsupervised classification method was used to classify the NDVI time series images into NDVI clusters. Area fraction of the clusters by tiles were generated to analyse the relationship and variability of the NDVI clusters and the strata map(s). The Pearson Chi-square( $\chi^2$ ) test was employed to analyse the relationships between the NDVI clusters and the strata map. The result revealed that all strata maps had a significant relationship with the NDVI clusters at a 5% significance level. A one-way ANOVA was used to analyse the variability of the area fraction of crop field NDVI clusters and the strata map. The result showed general within and between strata differences, such that the administrative unit (woreda) map had the lowest within strata difference of the others. Finally, maximum variability (counts of significantly different pairs of strata) between strata was determined according to the statistical significance in area fraction of crop field clusters at a 5% significance value using ANOVA, Tukey-HSD, and Hochberg's GT2 post hoc test analysis of pairwise comparison. The statistical analysis result demonstrated that 60% of the CPS zone pairs of strata had a significant difference, 38% of the LH zone pairs of strata had a significant difference, and 51% of the administrative unit (woreda) map had a significant difference. The overall analysis result of ANOVA indicated that the large significant differences in the CPS zone map consider spatial variability between strata that can identify uniform areas and help to generalize large areas. On the other hand, the administrative unit map had minimum variability within strata than the others, which indicates the administrative unit map provides samples from a small area that can accurately represent the area. To reduce uncertainty in crop area estimation, area stratification is supposed to be representative, relatively uniform within and generally different from its neighbours. The findings support integrating the CPS zone map and the administrative unit map can provide a stratification method for fragmented landscapes like Ethiopia, reducing the uncertainty in crop area estimation.

**Keywords:** Stratification Map, Tiles, NDVI, ISODATA, Crop Field, Planet, Cluster Variability

## ACKNOWLEDGEMENTS

First, I would take this opportunity to thank the Netherlands Government and the Orange Knowledge Programmed (OKP) for the scholarship, which covers all financial costs and allows me to focus only on academics.

I want to express my heartfelt gratitude to my first supervisor, Dr. Kees de bie, for his continuous support, patience, guidance, and encouragement in every aspect of the research. He also gave me a valuable comment and helped me find ways to think critically. This research would not have been possible without his guidance and support.

I would like to thank my second supervisor, Dr. Claudia Paris, for her comments, support, and valuable input in this research. Additionally, I would also like to thank Dr. Michael Marshall, for his valuable comments on my proposal document.

Finally, I would like to thank my family for your support during these two years of my academic journey.

# TABLE OF CONTENTS

---

1.	Introduction.....	1
1.1.	Background and Motivation.....	1
1.2.	Problem Statement.....	3
1.3.	Research objectives .....	4
2.	Study Area and Dataset.....	6
2.1.	Study Area.....	6
2.2.	Datasets .....	8
3.	Method .....	11
3.1.	Creating Sampling Scheme .....	13
3.2.	NDVI Data Computation from Planet-Imagery .....	15
3.3.	Time Series NDVI Data Clustering.....	16
3.4.	Analysing the Relationship between Counts of Tiles by Tile Group and the Strata map.....	17
3.5.	Analysing the Variability of Area Fractions of the Crop field Clusters by Tile and Strata Map .....	18
4.	Result .....	19
4.1.	Sampling Scheme.....	19
4.2.	NDVI data Computation from Planets-Imagery.....	22
4.3.	Time Series NDVI Data clustering .....	23
4.4.	Analyzing the Relationship Between Counts of Tiles by Tile Group and the Strata Map.....	27
4.5.	Analyzing the Variability of NDVI Clusters (Crop Field) and Strata Map.....	32
5.	Discussion.....	34
5.1.	On Sampling Scheme.....	34
5.2.	On NDVI Data Computation .....	34
5.3.	On NDVI Data clustering .....	35
5.4.	On The Relationship Between Counts of Tiles by Tile Group and the Strata Maps .....	35
5.5.	On The Variability of Crop field Clusters and Strata Map .....	36
6.	Conclusion and Recommendation.....	38
	LIST OF REFERENCES.....	39
	ANNEXES.....	43

## LIST OF FIGURES

---

Figure 1: Map of the study area .....	7
Figure 2: Administrative unit map (enumeration areas) of Oromia region, (CSA, 2015).....	8
Figure 3: Oromia Region Livelihood zone map source; USAID, 2010.....	9
Figure 4: CPS zone map of Oromia region (produced by Mohammed et.al (2020) .....	10
Figure 5: Flow chart of the research method.....	12
Figure 6: Spatial distribution of sample polygons and tiles of the study area. Black area denotes polygons, and small red boxes within polygons denote tiles.....	14
Figure 7: Example of the organized image of the tiles (all images in one file) with represented stratification map zones/woredas(strata).....	14
Figure 8: Layer Re-stacking into 9-12, 1-12, 1-4. ....	16
Figure 9: The four hundred candidate polygons for sampling .....	19
Figure 10: The time series NDVI image of a single tile of the study area (2021).....	22
Figure 11: Examples of the organized (all in one file) NDVI image of the tiles rows indicates the sampled polygon and columns are repeats of tiles in the. ....	22
Figure 12: The 25 NDVI clusters of the tiles. ....	23
Figure 13: The 25 NDVI clusters grouped based on their similar characteristics, profiles (lines) denote the clusters, titles named groups denote the cluster after inspecting the GE image.....	24
Figure 14: NDVI Clusters class and its represented features on GE image; Both the green and blue arrow indicate crop fields. ....	25
Figure 15: NDVI cluster and its represented features on GE image; bare land indicated by red arrow and yellow arrow indicates fields without crop (uncultivated land).....	26

## LIST OF TABLES

---

Table 1: The total and selected number of woredas and zones(strata) of stratification maps and total number of selected tiles. ....	20
Table 2: Counts of strata of the three strata maps in the selected polygons, column labels indicate CPS zone, the row labels indicate the LH zone code (Bold) and woredas. ....	21
Table 3: Examples of the area fraction of the clusters by tiles according to the strata maps (strata maps versus NDVI clusters, representing area fraction). ....	27
Table 4: Grouped tiles (10) versus clusters (25) represents average area fraction in percent, the area fractions with crops filed highlighted in green. ....	28
Table 5: Crosstabulation table of grouped tiles (10) versus Strata of the CPS zone map representing counts of tiles. ....	29
Table 6: Crosstabulation table of grouped tiles (10) versus Strata of the LH zone map representing counts of tiles. ....	30
Table 7: Crosstabulation table of grouped tiles (10) versus strata of the woreda map representing counts of tiles. ....	31
Table 8: The three individual Chi-Square tests result to examine if counts of tiles are significantly related to the stratification map. ....	32
Table 9: The within and between groups (strata) difference in the crop field clusters of the three strata maps. ....	33



# 1. INTRODUCTION

## 1.1. Background and Motivation

Nowadays, agriculture is continuing to be a core instrument for sustainable development and poverty reduction. About 75% of the population of developing countries lives in rural areas. Nearly all of them depend on agriculture to secure the necessities of life (World Bank, 2008). Agriculture plays a vital role within the economies of developing countries in offering the most supply of food, revenue, and work; and its development is of central importance to meeting food security, poverty reduction, and general sustainable development (UDA Consulting, 2018). According to the World Bank (2010), decisions regarding the aid and development of agriculture and its investment require agricultural statistics relating to land usage, agricultural production, current social and economic conditions that the farmers confront, and other related concerns. Currently, also because of the requirement to recognize the effects of rising population and income on food security, weather change, and to meet the Sustainable Development Goals 2 (SDGs), particularly "*SDG indicator 2.4.1 (proportion of agricultural area under productive and sustainable agriculture)*," there is a growing request/demand for correct, reliable, and frequent agricultural data (FAO, 2018). Agricultural statistics are essential for sound policy decisions, ensuring the efficient operation of the market and promoting investment (FAO, 2015). However, high uncertainty in agricultural statistics can impede effective policymaking and planning.

Agricultural statistics, particularly crop production data (crop area  $\times$  yield), play a significant role in resource planning and allocation for the agriculture sector growth (Sud et al., 2017). In addition, it is used as an economic indicator for early warning signals, foreign aid assessment for the requirements of food import, market trends and analysis, and trade policy and exporter support (Dahal et al., 2018). Planning and policymakers require reliable and timely information on crop area and yield data to formulate efficient agricultural policies and make important decisions about purchasing, storing, distributing, importing, exporting, and other related problems (Sud et al., 2017). Therefore, crop area estimation is a critical component of every country's agricultural statistics system since it is vital in assessing crop production (Sud et al., 2017).

The optimal method for assessing crop area is determined by various operational criteria, including land arrangement, crop field shape, crop type, cropping pattern, and available skills and resources (Sud et al., 2017). A solid assessment of crop areas in developing countries with persistent food deficiencies is fundamental for emergency support and the construction of proper market-based food security strategies (Husak et al., 2008).

In the Sub-Saharan African countries, farmlands are small (less than two hectares), dispersed, and complex (Eggen et al., 2016; Mohammed et al., 2020a), making estimates of crop area challenging and necessitating better quality maps (Eggen et al., 2016). Besides, high landscape fragmentation like elevation/soil type, inequitable land tenure, and the extreme variability of the environment over space and time and mixed cropping systems are also challenging for crop area estimates (Aguilar et al., 2018). In such regions, estimating crop areas requires higher quality data to identify the extent of small and fragmented farmlands. For being ready, for food supply problems, instability in agricultural market prices, and other related issues, reliable and appropriate agricultural information like crop area estimates on present and prediction food production is required. Thus, the most appropriate methods of crop area estimation will lead to better prediction and recognize and minimize the problem related to food insecurity.

There have been numerous techniques for estimating crop area over the years (Craig and Atkinson, 2013). The census and survey method of agricultural statistics collection is a traditional method that entails enumeration of the entire population and farms (Craig and Atkinson, 2013), then taking a sample of a farm area that needs sample households within the samples. Even though it is helpful as an information source for crop area estimation and food security assessments, this type of method is expensive and labour-intensive, which makes it challenging to acquire representative estimates over space and time (Craig and Atkinson, 2013), which means it can provide accurate but less representative data. Thus, it is difficult to obtain accurate estimates over large areas and meet the current demands for agricultural data and food security assessment in smallholder farming systems with such a method. According to a World Bank (2011) report, adequate agricultural data for monitoring agriculture is difficult to gather and report in developing countries due to financial constraints, a lack of labour, and insufficient statistical approaches.

Remote sensing image-based methods have been used to develop many methodologies for crop area estimation (Husak et al., 2008). Such as supervised classification (Natteshan and Suresh Kumar, 2020), area frame sampling and high-resolution imagery (Grace et al., 2012), Bayesian/fuzzy classification (Gorte and Stein, 1998), pixel count (Bauer et al., 1978; Gallego, 2004; Maselli et al., 2011; Shao et al., 2001) spectral unmixing, area frame sampling (Gallego, Francisco, 1999; Pradhan, 2001; Taylor et al., 2010) landscape stratification of MODIS time series (Vintrou et al., 2012). Recently more machine learning approaches have been applied for crop area mapping in smallholder farming. Such as cloud-based multitemporal ensemble classifiers (Aguilar et al., 2018) to create a map of the most common crops in west African smallholders and map agricultural field size globally using crowdsourcing (Lesiv et al., 2019).

A variety of image classification approaches have been applied to determine the agricultural land limit and estimate crop area using remote sensing imagery. The existence of mixed pixels, geolocation error, and the spatial relationship between training and test locations restrict the pixel count that is limited by classification accuracy (Husak et al., 2008). Additionally, classification and spectral unmixing algorithms are extremely sensitive to class variance and spectral band correlation (Husak et al., 2008). The majority of image classification algorithms depend entirely on the spectral information included in remote sensing images. Nevertheless, numerous studies have employed supplementary and geographical information to increase the accuracy of crop area demarcation and estimation (Recio et al., 2010). Although fine resolution remote sensed data and machine learning approaches provide potential solutions, they are plagued by logistical problems that are exacerbated in small holder farming areas (Mohammed et al., 2020a).

Area Frame Sampling (AFS) gives an alternative method to reduce the expense of conducting wide-field surveys using data collected (i.e., sample) at various scales to establish a relationship among sampling stages (Husak and Grace, 2016). It makes extensive area estimation using dense but geographically limited data and relations with coarse but broadly obtainable data. The dense and geographically limited data might be obtained from fieldwork, farmer interviews, high-resolution imagery, and aerial photographs, whereas the coarser and broadly available data acquired from low-resolution satellite imagery, census results, and classified satellite data (Husak and Grace, 2016). The crop area estimation of this method would be over a wide area. AFS uses a set of geographic units rather than an administrative unit that can be points, transects/segments of land to construct primary sampling units, and secondary sampling units (FAO, 2017); however, the user community chose to adopt administrative units as strata especially in developing countries. Gallego (2004) stated that AFS has its own limitations in terms of position and interpretation inaccuracies, but some applications utilizing impartial estimation methods (ground survey or high-resolution aerial pictures) demonstrate little variance from real crop inventories. Besides, the AFS approach faces the same difficulties as the traditional methods, as both are expensive, time-consuming,

and labour-intensive. Furthermore, in the vast majority of situations, the samples are insufficient to be generalized over huge areas (Marshall et al., 2011).

The main sources of agricultural statistics for many developing countries are the agricultural census conducted every ten years, which is not adequate to analyse the present conditions (FAO, 2015). Such applications require current and correct data that might be given correctly through sampling. AFS and remote sensing data help to stratify the agricultural land for sampling and then enable to estimate of the crop area. Some of the developing countries are now using different area stratification methods for crop area estimates, such as administrative unit strata map (enumeration areas), livelihood zones strata map, crop production systems zones strata map, agroecological zones, etc. However, different factors determine the uncertainty of crop area estimates in smallholder farming areas. The main factors are how to define the strata of an area stratification methods, and the relationship between remote sensing data and crop areas are different according to crop production areas (a model for estimating crop area in one area will not work for another area). In addition, it is difficult to know which stratification works best since each method have its own advantages and limitation.

## 1.2. Problem Statement

Many developing countries are affected by the early warning system of food security because of inadequate agricultural data, like crop area estimates, which are used to determine food production (FAO, 2015). In sub-Saharan Africa, the problem of food security is occurring more frequently, particularly the problem is acute in areas with smallholder farming regions and fragmented landscapes like Ethiopia. Even though there have been various types of crop area estimates, the uncertainty of each type of crop area estimate varies with the topographical nature of the areas and other related factors. In areas with regular, wide, and well-delineated fields, remote sensed based estimates of field fraction have been proven at pixel or parcel level (Han et al., 2012), whereas estimating crop areas is difficult, especially in fragmented and smallholder farming areas (Fritz et al., 2019). Therefore, area stratification might be based on areas that do not properly represent the areas of actual crop fields, which leads to the biased representation of area stratification for of crop area statistics in smallholding farming areas. As an estimate of crop area is essential for an early warning system of food security, the estimate should be accurate and reliable. To get an accurate and reliable crop area estimation, identifying the most appropriate<sup>1</sup> area stratification method is very important, which in turn reduces the problem of food security that is caused by the lack of agricultural data such as crop area estimation.

This study focused on the comparison of only the three area stratification methods which have been used by the local government and non-governmental/aid organizations in East Africa, especially in Ethiopia. These are area stratification with an administrative unit map (Enumeration areas) (Central Statistics Agency (CSA), 2015), livelihood zones map (FEWSNET, 2021) and crop production system zones map (de Bie et al., 2011).

The comparison of the strata maps was performed using high resolution Normalized Differential Vegetation Index (NDVI) data from Planet imagery (PLANET, 2021) as an indicator. The use of remote sensing data can support monitoring of crop growth by giving accurate and timely information on vegetation development and phenology (Veloso et al., 2017). The NDVI time series is one of the most well-known methods for crop monitoring and spatial distribution of crops (Veloso et al., 2017). For the past few years, high-resolution imagery such as Landsat-8 and Sentinel-2 has become the primary source for determining crop areas having 30 m and 10 m spatial resolution, respectively (Werner et al., 2019). For

---

<sup>1</sup> **The most appropriate** area stratification method refers to a stratification method that is representative, relatively uniform within and generally different from the neighbouring strata to reduce the uncertainty of crop area estimation relative to the others.

this study, the NDVI data was obtained from Planet imagery, which provides very high-resolution images (i.e., 4.77m) and has been freely available since September 2020 for only tropical areas, offering a time series of images. The temporal resolution (daily based) of Planet imagery helps to deal with cloud coverage in the study areas better than Landsat-8 and Sentinel-2 images.

In this study, statistical analysis (i.e., chi-square test and Analysis of Variance (ANOVA)) was used to determine the relationship and variability of NDVI clusters according to the strata maps. The Pearson chi-square test statistics were used to test the relationship between the row and column variables (Shih and Fay, 2017). An ANOVA test was used to analyse the variability of NDVI clusters according to the strata maps. In this context, the relationships of NDVI clusters (converted into area fractions and then grouped by tiles) according to the strata maps were analysed. The variability of the area fraction of clusters was analysed using the pairs of zones/woredas of the strata maps.

This study aims to compare and identify the most appropriate area stratification method in smallholder farming areas for crop area estimation by comparing crop production systems zones, livelihoods zones, and administrative unit maps (enumeration areas) using very high-resolution monthly NDVI time series data as an indicator.

### **1.3. Research objectives**

#### **1.3.1. General Objective**

The aim of this study is to compare three existing stratification methods for Oromia, Ethiopia, to capture and monitor crop area statistics. The stratification must be significantly related to differences in land cover types. If so, secondly, the best<sup>2</sup> one must have a minimum variability within strata and a maximum variability between strata as related to arable crop fields present.

#### **1.3.2. Specific Objectives**

1. To prepare a sampling scheme of the monthly Planet imagery that specifies where clusters of 1km x 1km image samples (tiles) must be taken.
2. To extract the time series NDVI (12 month) of each 1km x 1km image-sample (tile).
3. To classify the time series NDVI of all 1km<sup>2</sup> sampled tiles into the NDVI clusters.
4. To analyse the relationship between counts of tiles by tiles-groups and the strata maps [generalized strata map analysis]
5. To analyse the variability of area fraction of crop fields cluster by tiles versus strata maps [detailed strata map analysis]

#### **1.3.3. Research question and Hypothesis**

RQ-1: Is the counts of tiles by tile groups significantly related at 5% significance level (sig. level) with strata maps?

H0 1: There is no significance relationship at 5% between counts of tiles by tile-group and the strata map.

H1 1: There is a significance relationship at 5 % between counts of tiles by tile-group and the strata maps

RQ-2: which strata map has more counts of significantly different pairs of strata in area fraction of crop field clusters at 5% sig. value?

H0 2: The counts (percentage) of significantly different pairs of strata in area fraction of crop field clusters at 5% sig. value is lower for the CPS zone strata map relative to the others.

---

<sup>2</sup> **Best** refers here to a stratification method that is representative, relatively uniform within, and generally different from the neighbouring strata to reduce the uncertainty of crop area estimation relative to the others.

H1 2: The counts (percentages) of significantly different pairs of strata in area fraction of crop field clusters at 5% sig. value is the highest for CPS zone strata map relative to the other.

## 2. STUDY AREA AND DATASET

### 2.1. Study Area

Ethiopia's economy is based on agriculture, which generates for 40% of GDP, 80% of exports, and over 75% of the country's employment (USAID, 2020). Ethiopia is among the world's most food insecure and famine-prone countries (Mohamed, 2017), and in the 2021 global hunger index (GHI), Ethiopia is ranked 90<sup>th</sup> out of 116 countries with a score of 24.1, indicating that the country suffers from severe hunger (GHI, 2021). The direct contributor to food insecurity in rural regions of Ethiopia is drought and rainfall unpredictability because it has a significant impact on food availability and access. However, several reasons contribute to Ethiopia's food insecurity, including environmental deterioration, population pressure, poor agricultural policy and technology, low educational background (WFP and CSA, 2019), and internal unrest and conflicts more recently. It is crucial to assess the provision or supply of food in Ethiopia at each country-wide and regional level to increase systematic understanding, minimize malnutrition, and alleviate food insecurity. Creating yearly indicators of local food production assists in the identification of food-insecure groups and families (Barrett, 2010). However, concerns about data accessibility and quality impede a valid seasonal or yearly assessment of local food supply (Husak and Grace, 2016).

Therefore, the Oromia region was selected as a study area. Even though Oromia is agriculturally productive, food security is a severe problem, particularly in the southern and eastern parts of the region. In addition to the various climatic zones of the Oromia region, as described by Eggen et al. (2016), farmlands are small, dispersed, and complex, and various agricultural statistics-related studies have been conducted, for example, the three selected stratification methods. The government and aid organizations use different methods for agricultural statistics collection. This makes the region an exciting place to compare the three stratification methods for agricultural statistics collection. Details of the study area are presented in the following sub section.

#### 2.1.1. Geography

The Oromia region is one of the eleventh regional states of Ethiopia, which covers parts of the central, western, southern, and south-eastern parts of the country. The country's capital city, Addis Ababa, is also found in the Oromia region. The region extends between 03°00'-10°34' N latitude and 34°07'-42°55' E longitude (Figure 1). The total area of the Oromia region is about 353,690 Km<sup>2</sup> (the largest in Ethiopia) and is administratively divided into 17 zones and subdivided into 249 woredas (i.e., districts)(Adugna Aynalem, 2021) . Oromia is Ethiopia's most agriculturally productive area, with smallholder farming dominating. The great east African rift valley divided the region into the west and east with three different climatic zones (i.e., tropical, sub-tropical, and temperate zones) (Mohammed et al., 2020).

#### 2.1.2. Topography and Climate

The topography of the Oromia region involves high and steep mountain ranges, undulating highlands, wide-angle valleys and river valleys, and rolling plains. The region's altitude rises from the lowest point, less than 500m above sea level, to the highest mountain, at 4607 m (Mount Batu). In addition to its distinct relief features, the region is characterized by diverse and pleasant climatic conditions and a large natural resource base (Adugna Aynalem, 2021). There are three major climatic types in the region, a dry climate, a tropical wet climate, and a temperate wet climate. A dry climate is characterized by poor vegetation, an annual mean temperature ranging from 27 to 39°C and an annual rainfall of fewer than 450 millimetres. A tropical wet climate has a yearly mean temperature ranging from 18 °C to 27 °C and a mean annual rainfall range of 410-820 mm with significant variation from year to year. The Oromia Highlands

has a temperate climate with moderate temperatures (the coolest month's mean temperature is less than 18°C) and plenty of precipitation, 1200-2000 mm (Adugna Aynalem, 2021).

### 2.1.3. Population and Agricultural activity

The population of the Oromia region was projected to be about 38 million in 2018 (35.1 % of the country's population), and only about four million people live in urban areas (Government of Ethiopia, 2018). Agriculture is the region's main economic activity, and Oromia is the country's largest crop-producing area, owing to relatively ample rainfall, suitable soils, and other agricultural potentials (Adugna Aynalem, 2021). Ethiopia has two main crop growing seasons: the Meher and Belg. The Meher (primary season, accounting for 96 percent of overall production) and the Belg (secondary season) are the two growing seasons in Ethiopia mainly for smallholders (Alemayehu et al., 2012). The Meher season rain starts in June and ends in October, and Teff, wheat, maize, and sorghum are the principal crops cultivated in the Oromia region during Meher. The Belg season is shorter and gets less rain on average than the other seasons. Its season starts in February and ends in April or May (FAO and WFP, 2007). Potatoes and yams are the main crops cultivated during this season (Husak et al., 2008).

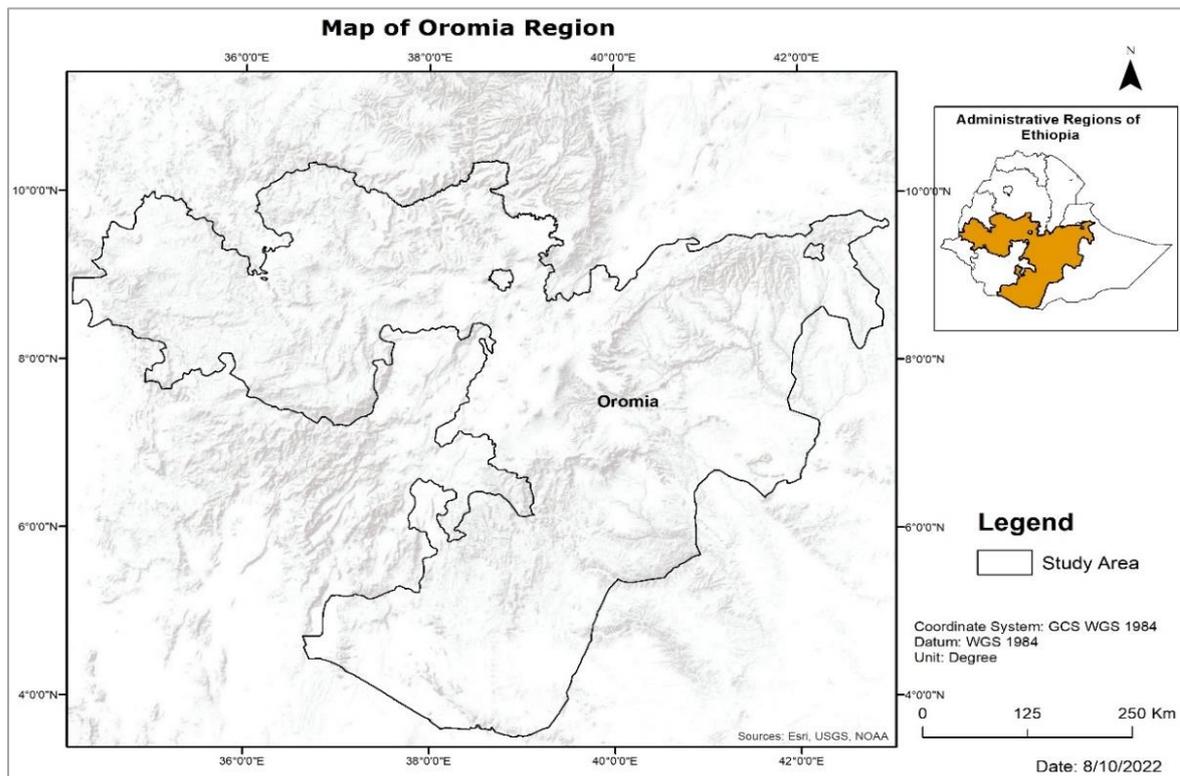


Figure 1: Map of the study area

## 2.2. Datasets

In this research, different data sources have been used to compare the three area stratification methods. The data included Planet imagery through the Google Earth Engine (GEE) platform, three area stratification methods (Crop Production System (CPS) zone, Livelihood (LH) zone, and administrative unit (woredas)) map and Google Earth (GE) imagery.

### 2.2.1. Administrative unit map (woredas)

The agricultural census of administrative unit map<sup>3</sup> (enumeration areas) is a non-overlapping subsection of a country for which agricultural census information will be accessible, and it's appropriate mapping of enumeration areas would maintain adequate coverage and prevent duplication and omission (FAO, 2021). A list of enumeration areas along with the projected number of holdings in each enumeration area makes up an agricultural census frame. The agricultural census frame can also be constructed using national population and housing census data and maps in various ways (FAO, 2021). Area stratification with enumeration areas utilizing a record of enumeration areas which was acquired from a census of population and housing a collection of sampled enumeration areas selected based on the planned sample design for survey. A fresh list of households in each selected enumeration area was created from selected enumeration areas. A list of agricultural households in each enumeration area was utilized as a sampling frame, and the agricultural households were ultimately chosen as data sources (Central Statistics Agency of Ethiopia, 2015). For this study Oromia's Administrative unit map was used. Figure 2 shows the administrative unit map (enumeration areas) of Oromia.

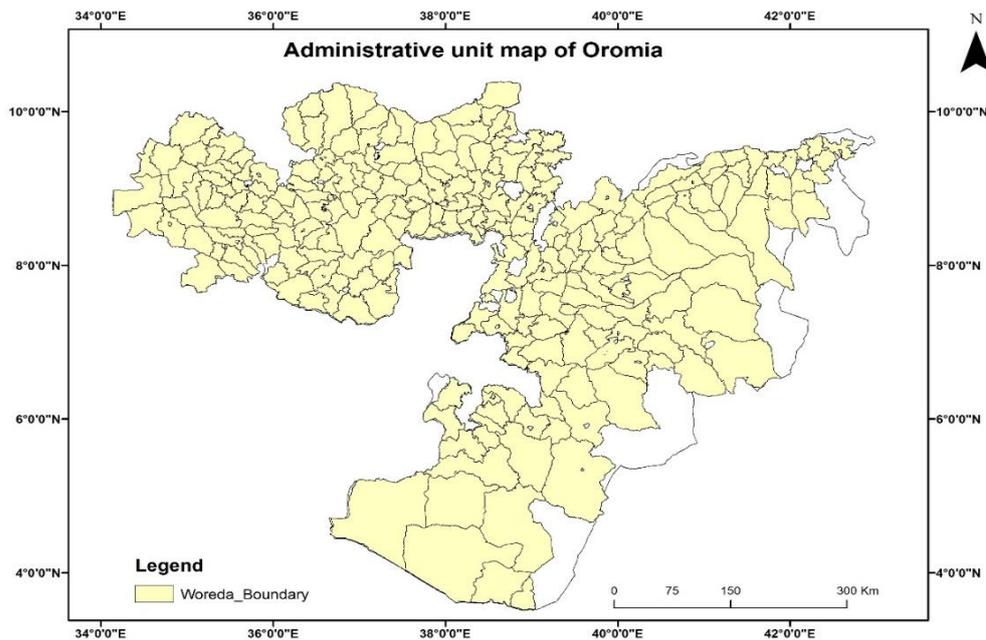


Figure 2: Administrative unit map (enumeration areas) of Oromia region, (CSA, 2015)

<sup>3</sup> In this research, the terms administrative unit map, enumeration areas and woreda map used interchangeably to refer to administrative unit strata map

**2.2.2. Livelihood zone (LH)**

A livelihood zone is an area where individuals follow a similar livelihood pattern encompassing access to food and income sources (which means people farm the same crops or rear the same kinds of animals), as well as market opportunities (FEWSNET, 2021). It is essential to provide a sampling frame/stratification for crop area estimation and the geographic location of the livelihood system to help with food security assessment and aid targeting (FEWSNET, 2021). Marshall et al. (2011) used the livelihood zone maps with elevation and NDVI as input for stratification of the area in Niger, Africa. They discussed that the livelihood zone map is a significant indicator of crop area, and large sampling schemes that can be analyzed using high-resolution images and deliver 'ground truth' across a vast area can be designed. Thus, using data like livelihood zones, elevation, and NDVI, the associations created can subsequently be generalized over entire countries. Such a method improves the usage of an area-frame sampling strategy for early warning networks in which image analysis gives broad-area crop percentages that may be utilized to reliably develop statistical models connecting physical and social data to crop area (Marshall et al., 2011). For this study, the 2010 Oromia Livelihood Zone map was used. Figure 3 shows the livelihood zone map of Oromia.

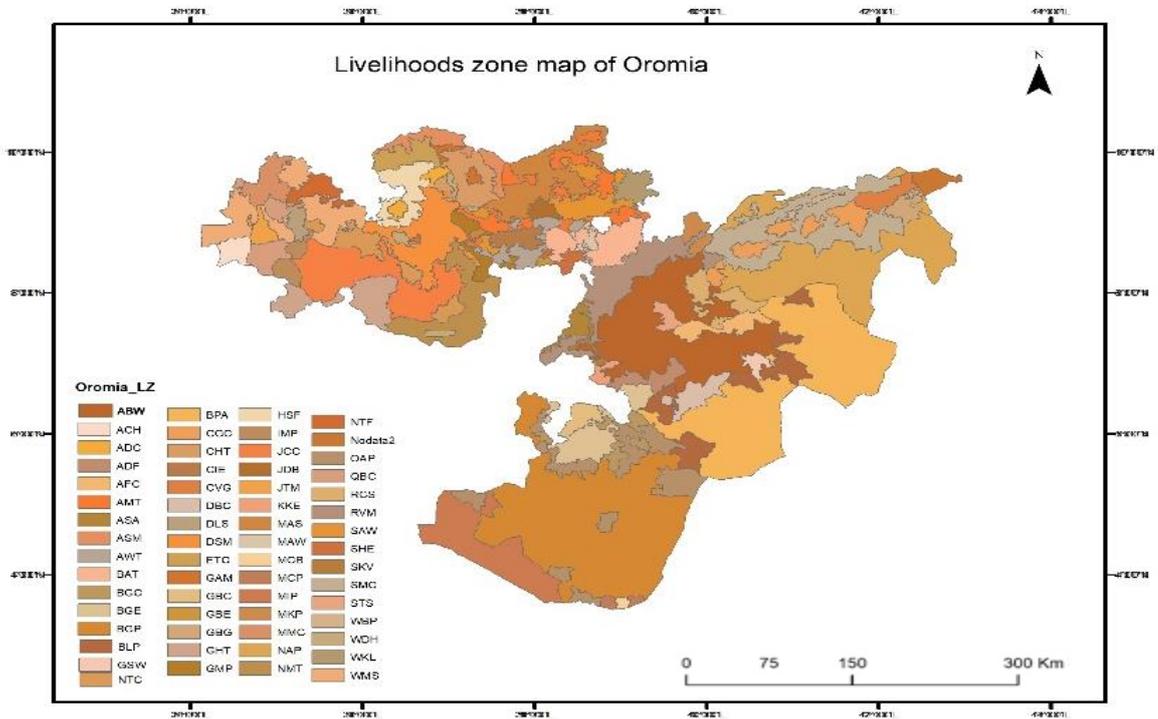


Figure 3: Oromia Region Livelihood zone map source; USAID, 2010

**2.2.3. Crop Production System Zone (CPS zone)**

Crop area can also be estimated using stratification in a combined agricultural census and remote sensing technique (de Bie et al., 2011). They used 1km spatial resolution and ten-day SPOT time-series images to blend with stated crop area statistics to create a crop area map in Nizamabad (India) and Andalusia (Spain). ISODATA, an unsupervised classification method, together with a ten-day SPOT Proba V 1 km NDVI time series, was used to stratify the terrain (land) into crop production system zones (CPS). This stratification method entails clustering the terrain into relatively homogeneous sections with comparable temporal patterns in the NDVI on a pixel-by-pixel basis (Mohammed et al., 2020). The CPS zone of

Oromia produced by Mohammed et al. (2020) was used. Figure 4 shows the CPS zone map field fraction at 1 km in the Oromia region.

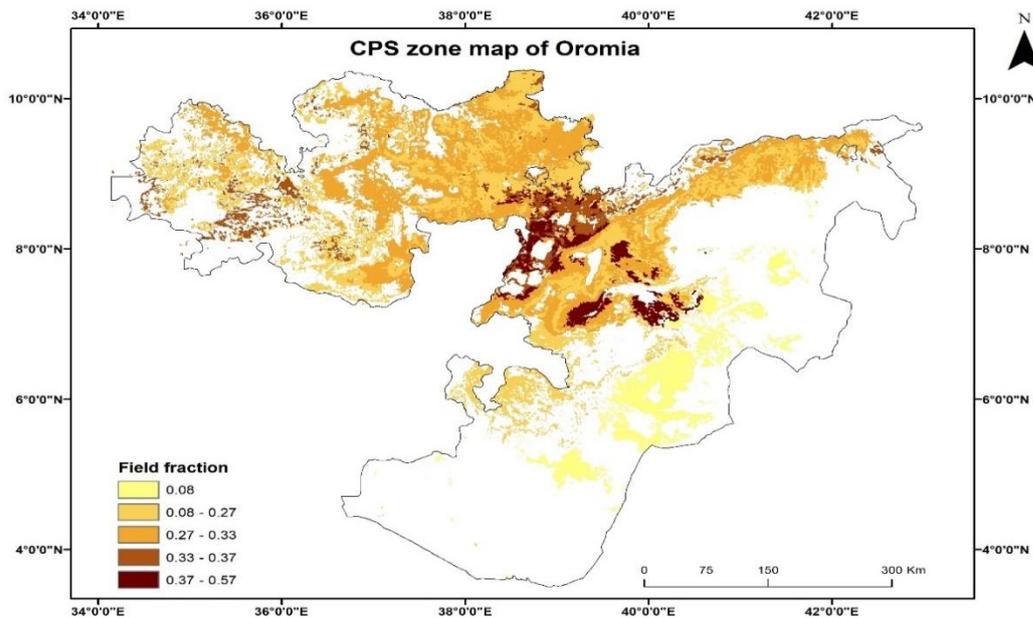


Figure 4: CPS zone map of Oromia region (produced by Mohammed et.al (2020))

#### 2.2.4. Planet Imagery (NICFI-funded planet tropical base maps)

Planet delivers daily and high-resolution (4.77m) satellite data for various applications and changes the earth observation business by making the highest frequency satellite data commercially available. The Norwegian Ministry of Climate and Environment granted Kongsberg Satellite Services (KSAT), along with significant other Planet and Airbus, an international agreement in September 2020 to offer complete access to high-resolution satellite surveillance of the tropics (Pandey et al., 2021).

Planet's satellite has four spectral bands: blue (455-515 nm), green (500-590 nm), red (590-670 nm), and near-infrared (780-860 nm) which are used to acquire images on a daily revisit time. Tropical Mosaics can now be accessed and downloaded with normalized (effects of atmosphere and sensor characteristics minimised). The access of Planet images, which is funded by Norway's International Climate and Forest Initiative (NICFI) has been free since September 2020 and can be integrate with platforms such as Google Earth Engine (Planet Team, 2017). This study uses the monthly composite images of 2021 using GEE platforms.

### 3. METHOD

The proposed method uses high-resolution Planet monthly composite images to compare land cover variability, mainly crop fields, between three stratification maps using time series of NDVI profiles associated to different land cover classes. The goal of the method is to identify the most appropriate area stratification method for crop area statistics in fragmented landscapes/smallholder farming area.

The flowchart of the method developed is shown in Figure 5 and is broadly defined as; i) Creating sampling scheme; ii) NDVI data computation from Planet imagery; iii) NDVI-Time series data classification; iv) Analysing the relationship between counts of tiles by tile group versus strata map and v) Analysing the variability of the area fraction of crop field clusters by tile versus the strata maps. The details are described in the following sub-sections.

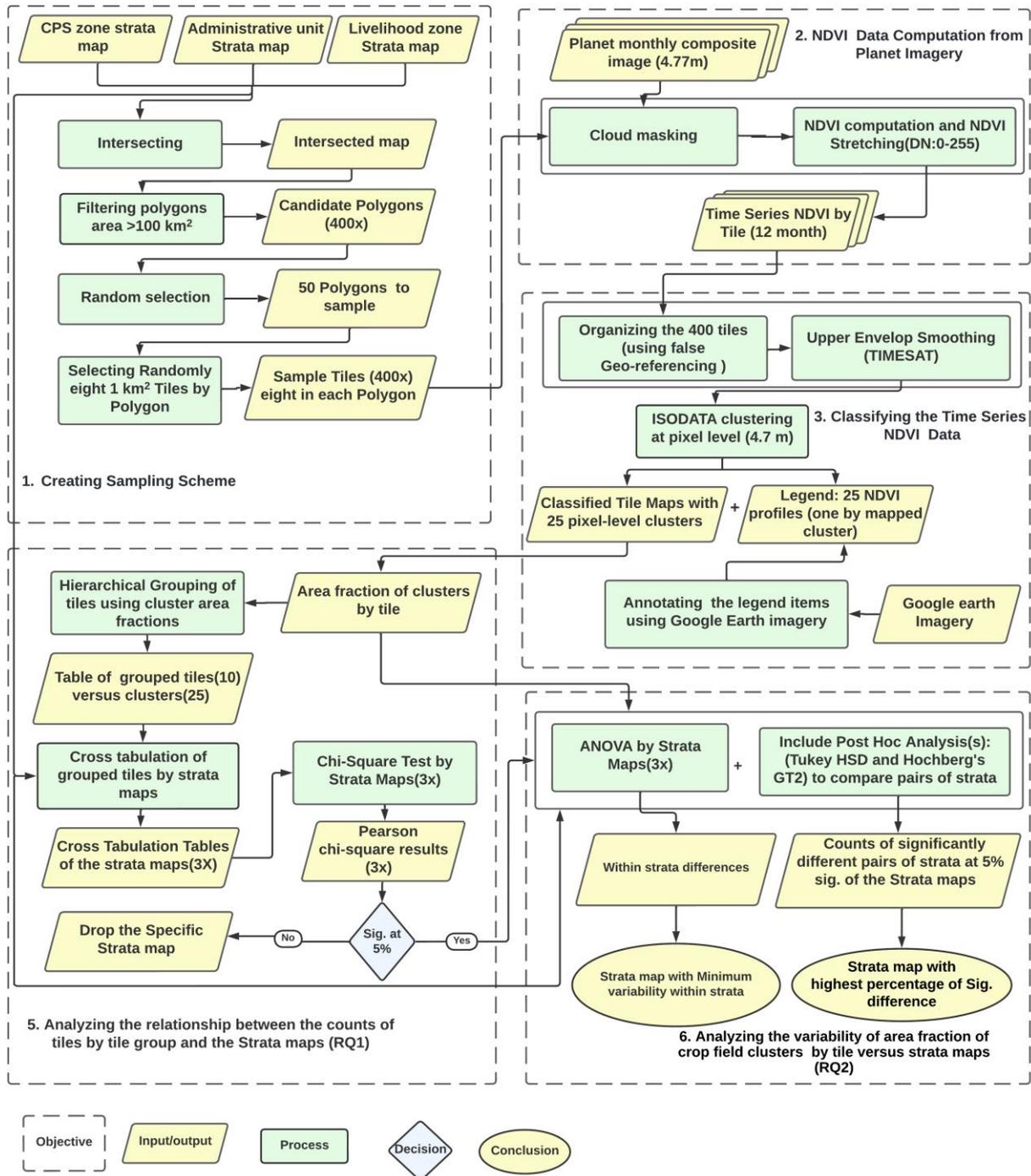


Figure 5: Flow chart of the research method

### **3.1. Creating Sampling Scheme**

A two-stage sampling scheme strategy was employed for representing the land cover variability of the study area. The considered sampling strategy is resembles to the one presented by Wang et al., 2018, who performed a two-stage sampling to estimate crop acreage in Mengcheng County, Anhui Province, China.

The area is divided first into primary sampling units (PSU), or clusters composed of secondary sampling units (SSU) such as farm listing, land unit segments, or points (Gallego, 2015). In some circumstances, two-stage sampling (also known as twofold sampling) might give an alternative to complete stratification if the surveyed area is too broad (Gallego, 2004). An extensive sample is chosen using a simple procedure in the first stage. Gallego, Francisco, (1999) indicated that this method was applied in central and eastern Europe using 40 x 40 km as a sampling unit using SPOT.

#### **3.1.1. Intersection of the three strata maps**

A geometric intersection of the three-stratification maps (Crop production system zone, livelihood zone, and administrative unit map) was computed to get polygons overlaying all three maps. After intersecting the maps, polygons or portions of polygons that represent all three maps were generated.

#### **3.1.2. Filtering candidate polygons**

In the developed method, after intersecting the three stratification maps, polygons with an area greater than 100 km<sup>2</sup> were selected as candidate polygons. While performing intersection, very small portions of areas or pixel-level areas can be created as an output polygon. To avoid these small areas being selected as sample areas, filtering was done to remove polygons of less than 100 km<sup>2</sup>. Finally, 400 polygons were identified as candidate polygons to perform the two-stage sampling scheme.

#### **3.1.3. Random Selection of Fifty Polygons**

After selecting candidate polygons, a simple random sampling method was used to select the first set of homogenous sample polygons. This condition allows us to generate polygons having a unique value in the CPS zone, livelihood zone, and administrative unit map. Therefore, 50 polygons were randomly selected as PSUs among the candidate polygons.

#### **3.1.4. Selecting Randomly Eight 1 km<sup>2</sup> Tiles in each polygon**

A variety of sampling methods might be used as a secondary sampling stage involving simple random sampling and other types of systematic sampling (Gallego, 2015). In the proposed method, the secondary sampling unit (SSU) is randomly selected by generating 400 tiles, eight per polygon, having 1km x 1km size. The spatial distribution of the PSUs and SSUs in the study area is shown in Figure 6. Finally, the tiles image map model was changed (artificially georeferenced) or organized (mosaicked) to make the processing and analysis work easy. Figure 7 shows parts of the organized images of the tiles and represents stratification maps and zones and woredas. For each polygon having the specific value of CPS zone, livelihood zone, and woredas, we have the eight tiles having the size of 1km x 1km.

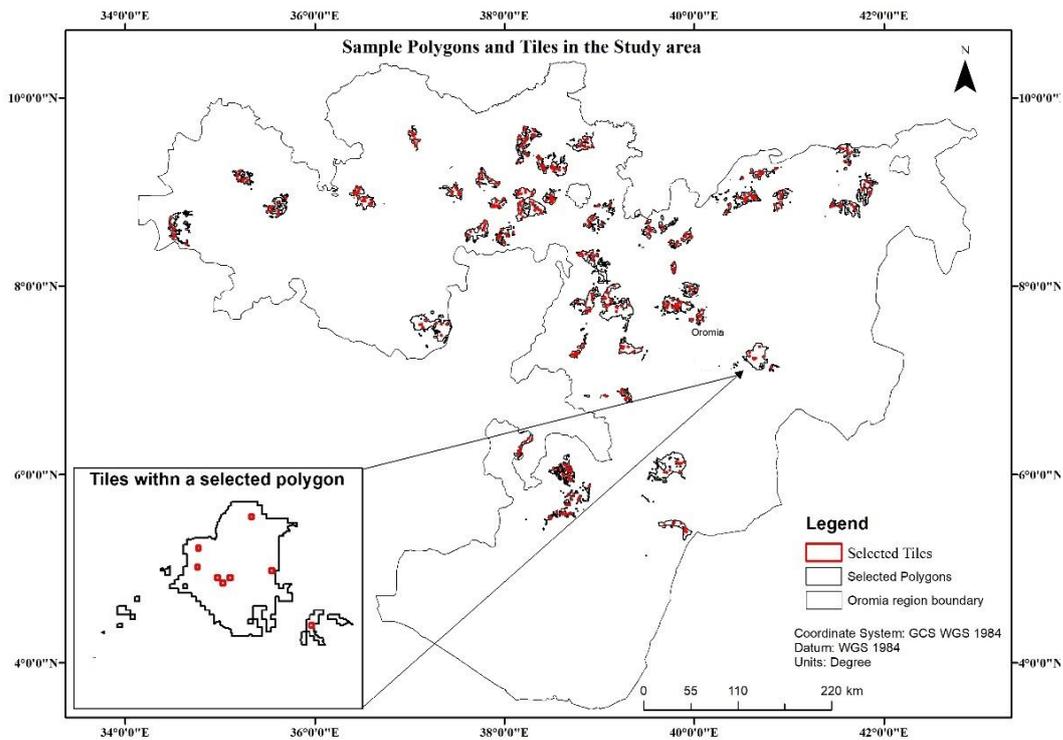


Figure 6: Spatial distribution of sample polygons and tiles of the study area. Black area denotes polygons, and small red boxes within polygons denote tiles.

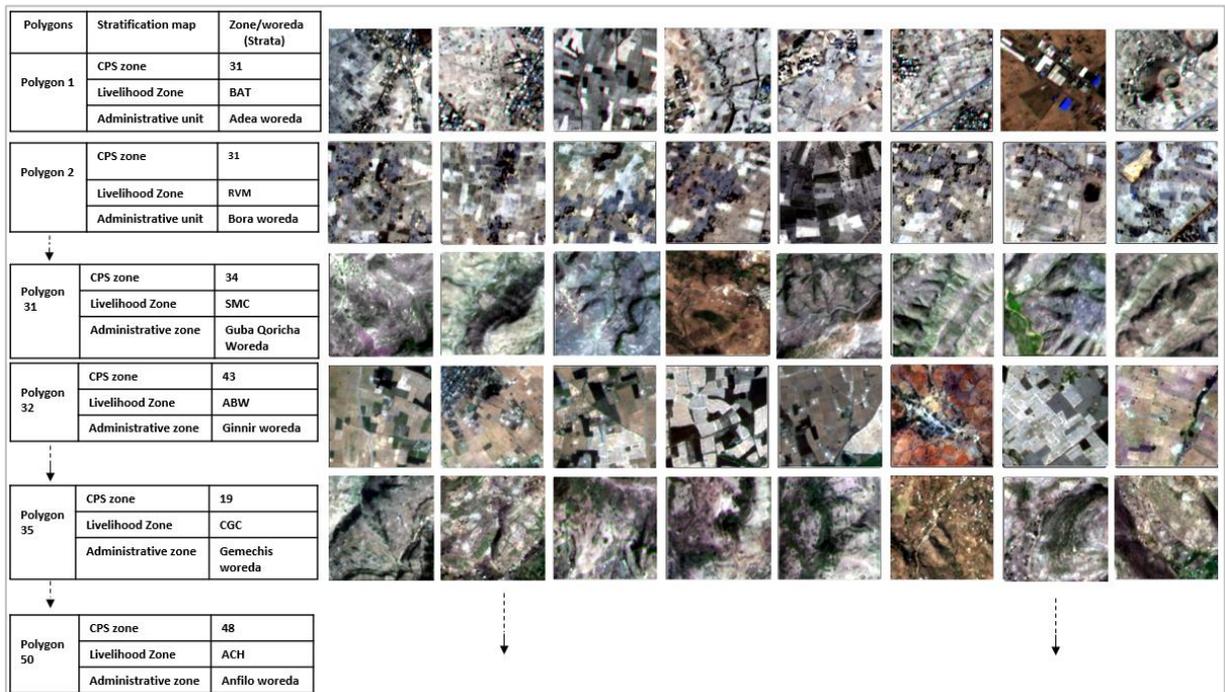


Figure 7: Example of the organized image of the tiles (all images in one file) with represented stratification map zones/woredas(strata).

### 3.2. NDVI Data Computation from Planet-Imagery

After identifying the sample areas, the NDVI computation started, considering the monthly composite of Planet images for 2021. Due to the heavy cloud cover, it was necessary to manually process each month individually to minimize the effects of the cloud cover. Therefore, in the cases of April, May, June, July, August, and September, it was necessary to perform a cloud masking procedure due to the cloud cover across large areas in Ethiopia present in these months. This problem is well known, and cloud masking is a critical pre-processing stage for data extraction activity of remote sensing images, which helps to estimate usable parts of the image (Malladi et al., 2018).

The NDVI data were computed in the Google Earth Engine (GEE) platform (Planet Team, 2017), which was used to assess the cloud cover and obtain the final NDVI output of the tiles.

#### 3.2.1. Cloud Masking

After the 12-month images of the study area were processed, cloud-covered and cloud-free months were identified by observing each month across the study area. Then, the necessary function for applying the cloud masking to the image collection was created by adapting the cloud masking algorithm directly on the GEE platform. The cloud masking was done by subtracting the median value of the green band of the time series image (12 months) from the median value of the green band of each cloud-covered month and setting some intensity values as a threshold. This intensity threshold was applied to filter clouds, like setting a cloud cover percentage, and the value is different according to the cloud cover level of each month.

#### 3.2.2. NDVI Computation and Stretching

Both NDVI processing and data stretching were computed for all months separately. After the NDVI processing, the image was rescaled (stretched) to digital numbers (DN) into 8-bit unsigned for ease of handling and better representation. It's worth noting that such rescaling does not result in data loss (Ali et al., 2013). Here, another essential step is also performed: setting zero for the missing value (clouds replaced by zero) that will be important in the time series computation, which can avoid removing/missing cloud-covered areas. Finally, the time series NDVI image was computed and exported only for the areas of the tiles. The computation was performed using the GEE platform (Planet Team, 2017).

NDVI is calculated by analysing satellite imagery to provide a measure of the “greenness”; as

$$NDVI = \frac{(R_{NIR} - R_{Red})}{(R_{NIR} + R_{Red})}$$

Where  $R_{NIR}$  represents the reflectance value of the Near-Infrared band,

$R_{Red}$  represents the reflectance value of the Red band.

The NDVI value ranges between -1 to 1, with higher values representing active and healthier vegetations (Meneses-Tovar, 2011). According to Meneses-Tovar 2011, NDVI values below 0.1 relate to water bodies and bare land, in contrast, higher values are signs of a high photosynthesis rate associated with crops (agriculture), shrubland, and forest.

NDVI of the high-resolution imagery is significant for identifying crop areas in smallholder areas because small-scale farming is more difficult to detect with coarse resolution imagery (Begue et al., 2014) since the mixed-pixel effect can significantly impact the spectral response of coarse or moderate resolution imagery (Pan et al., 2015). NDVI was used in this study as it is a reliable landcover indicator. Huang et al., 2021, stated that because of its spectral properties and relatively long history, NDVI had become the most widely used tool for vegetation assessment. This widespread application is due to the ease with which an

NDVI can be determined using any multispectral sensor with visible and Near-Infrared bands (Huang et al., 2021).

### 3.3. Time Series NDVI Data Clustering

#### 3.3.1. Upper Envelope Smoothing

Because of the limitations of optical sensors, it is not possible to observe the earth completely cloud-free or under ideal conditions every day. Remote sensing applications are affected by noise caused by unfavourable atmospheric conditions and changes in the sun's zenith angle throughout the year (Hird and McDermid, 2009). NDVI time series images are also hindered by prevalent noise resulting from such factors (Hird and McDermid, 2009). As a result, data smoothing or noise reduction is essential before further analysis.

After exporting the four hundred NDVI images, the image map model was changed (artificially georeferenced), and mosaics of the images were created using ERDAS IMAGINE to simplifying the analysis process; Figure 8 shows the organized (mosaicked) images. Then the images were stacked into 9-12, 1-12, and 1-4 to make them ready for smoothing/noise reduction. Originally the stacked layer was 1-12, which refers to the time series of the 12 months, then adding three months of the same year, the first three, and the last three months at the end and starting of the year. This was computed to have a sequence of NDVI profiles, as shown in Figure 4.

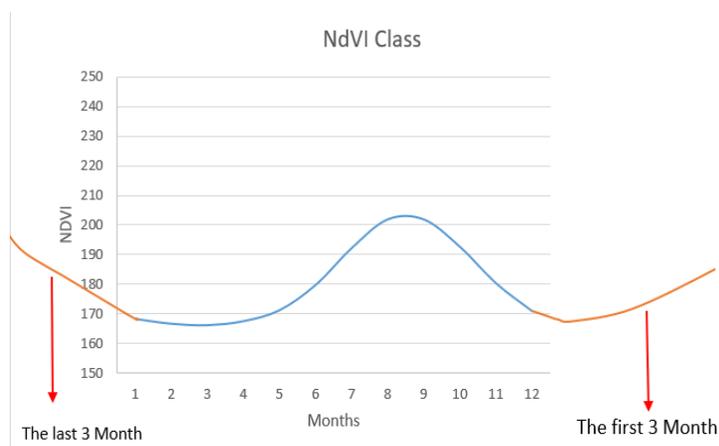


Figure 8: Layer Re-stacking into 9-12, 1-12, 1-4.

Then the TIMESAT was made in IDL-Envi software which makes the stacked image is to clean (de-haze and de-cloud) by pixel of the NDVI data. The TIMESAT method uses the “Savitzky-Golay” filters to smooth noisy signals (Savitzky and Golay, 1964), which use successive windows to calculate the regression function. The estimated value and the actual values are then compared. So, negative anomalies are fully removed, curves aim to connect the upper part of the readings, and positive spikes are nicely cut-off. The upper envelope filter replaces the NDVI value related to irregular and unexpected dips with interpolated values and follows good estimates.

#### 3.3.2. ISODATA Clustering at Pixel Level

After smoothing the time series of NDVI, the ISODATA unsupervised clustering method was employed to generate blends of land cover with similar surface responses. The algorithm uses clustering techniques to automatically cluster highly comparable pixels into groups (Kvamme et al., 2019). The ISODATA clustering approach, like the sequential method, employs spectral distance to repeatedly classify pixels,

modify parameters for each class, and classify again so that the spectral distance patterns in the data are preserved data emerges through time (ERDAS, 2003). Unsupervised clustering was used since the algorithm depends on limiting the participation of the users. ISODATA clustering algorithm generates a set of “*spectral classes*,” which can subsequently be given class labels and labelled as “*information class*” (Kvamme et al., 2019). Twenty-five clusters were produced to analyse the land cover, and the number of iterations was set as 50 with a convergence threshold of 1.

### 3.3.3. Annotating the legend Item using Google Earth Imagery

The 25 NDVI clusters that were generated using the ISODATA unsupervised clustering method were manually labelled through inspecting the Google Earth (GE) imagery and the NDVI profiles. A GE imagery was used to inspect the representation of the profile by displaying each cluster and its image to identify the represented features of the cluster. Then, the NDVI clusters were labelled with their feature representation. The annotation was also enabled to identify groups of clusters with similar spectral characteristics. Finally, table of area fraction of clusters by tiles that can be identified as five similar spectral characteristics group was generated.

## 3.4. Analysing the Relationship between Counts of Tiles by Tile Group and the Strata map

### 3.4.1. Hierarchical Grouping of Tiles using Cluster Area Fractions by Tile

In this step, hierarchical clustering was performed using the area fraction of the clusters by tile to analyse the relationships between the counts of tiles in the hierarchical group and the strata map, which can give a generalized analysis of the strata map. Hierarchical clustering approaches seek to identify relatively homogeneous groups of variables based on selected characteristics and detect clusters within a dataset to reduce clustering while maintaining the structure of the data (Gonçalves et al., 2008). It uses an algorithm that starts with each variable in a single cluster and merges the clusters until only one is left. Here the cluster analysis combines the tiles of the clusters into new groups based on some statistics methods (between-group linkage) and measures (squared Euclidian distance) to generate the average linkage between groups. Therefore, with a range of minimum five and maximum ten groups of average linkage between groups of tiles versus the clusters were generated using SPSS software and the maximum was used for analysis.

### 3.4.2. Cross Tabulation of Grouped Tiles by Strata Maps

The ten groups by tiles (average linkage between groups) were used to generate a cross tabulation table for each of the strata maps that shows the counts of tiles in the groups versus strata of the strata maps. Finally, the Pearson chi-square test was employed to analyse the significance relationships.

### 3.4.3. Chi-Square Test by Strata Map

The Chi-Square test was implemented to examine the relationship between counts of tiles and the strata maps. The Pearson chi-square ( $X^2$ ) test is widely used to test relationships between two categorical variables (Shih and Fay, 2017). It is expressed as the following formulas:

$$x_c^2 = \frac{\sum (O_i - E_i)^2}{E_i}$$

Where c is degrees of freedom

O is the observed value

E is the expected value

The Pearson chi-square test for the three strata maps were computed to analyse the relationship between the counts of the tiles and strata maps. In other words, the relationship between the identified land cover and the strata maps were analysed since counts of the tiles indicates the landcover.

### **3.5. Analysing the Variability of Area Fractions of the Crop field Clusters by Tile and Strata Map**

#### **3.5.1. Analysis of Variance (ANOVA) Test by Strata Map**

The ANOVA test was employed to investigate the variability of the area fraction of the clusters according to the strata maps. Analysis of variance assesses the significance of the difference between groups (Cilek and Uslu, 2022; Jackson and Ferguson, 1972). ANOVA determines whether there is a significance difference between groups, but it does not examine which group or groups may be responsible for these differences (Cilek and Uslu, 2022). Statistical post hoc analysis was performed to identify which groups caused a difference if there was a difference between groups (Cilek and Uslu, 2022).

A one-way ANOVA test was conducted for each strata map to analyse the significant difference between strata (zones/woredas). The ANOVA tests show significant differences according to all three maps at 5 % sig. value, but it indicates whether there is a significant difference. However, it is crucial to determine which strata map measures the heterogeneity of strata between and the homogeneity of strata within., a post-hoc analysis was performed with all NDVI clusters to identify which strata map pairs of strata have more significance differences. Finally, an ANOVA was checked with crop field clusters (groups) to identify the within and between groups differences of the strata maps specifically to crop fields land cover type.

#### **Post hoc Analysis**

A post hoc test is required to determine how the pairs of strata differ significantly(Cilek and Uslu, 2022). For the study, it is important to know the difference between the strata of the strata map. The post hoc result provides a multiple comparison table showing a significant difference between all pairs of strata, enabling us to determine the variability between strata. However, several post-hoc statistics can be used to determine the group/groups that make a difference; choosing the right ones requires making certain assumptions. Equality of group variances is a crucial consideration when using the post hoc analysis method to compare many pairs (Chen et al., 2020; Cilek and Uslu, 2022; Ochola et al., 2020). A Tukey Honestly Significance Difference (HSD) test was used with an equal number of sample sizes, and Hochberg's GT2 test was used with an unequal number of sample sizes (Ruxton and Beauchamp, 2008). Therefore, the Tukey HSD test was used for woredas because the numbers of tiles per woreda were almost equal, and Hochberg's GT2 test was used for CPS and LH zones because the numbers of tiles per zone were not equal.

## 4. RESULT

### 4.1. Sampling Scheme

#### 4.1.1. Intersected Map of the Three Strata Maps

The stratification maps, which classify the study area into different zones and woredas(strata), were intersected using Arc GIS software to produce an output map that represents the three strata maps. The administrative unit map classifies the study area into 245 woredas, the Crop Production System (CPS) zone map classifies it into 12 zones, and the livelihood (LH) zones map classifies it into 60 zones. After intersecting the three maps, more than 3000 polygons were created as an output map. Figure 9 shows the intersected map of the three stratification maps of the study area after filtering all polygons below 100 km<sup>2</sup> areas.

#### 4.1.2. Candidate Polygons(400x) for Sampling

After the intersection, many small polygons were created as an output that was difficult to sample. Thus, all polygons with less than 100 km<sup>2</sup> were filtered out, and 400 candidate polygons were identified. Figure 9 shows the 400 candidate polygons.

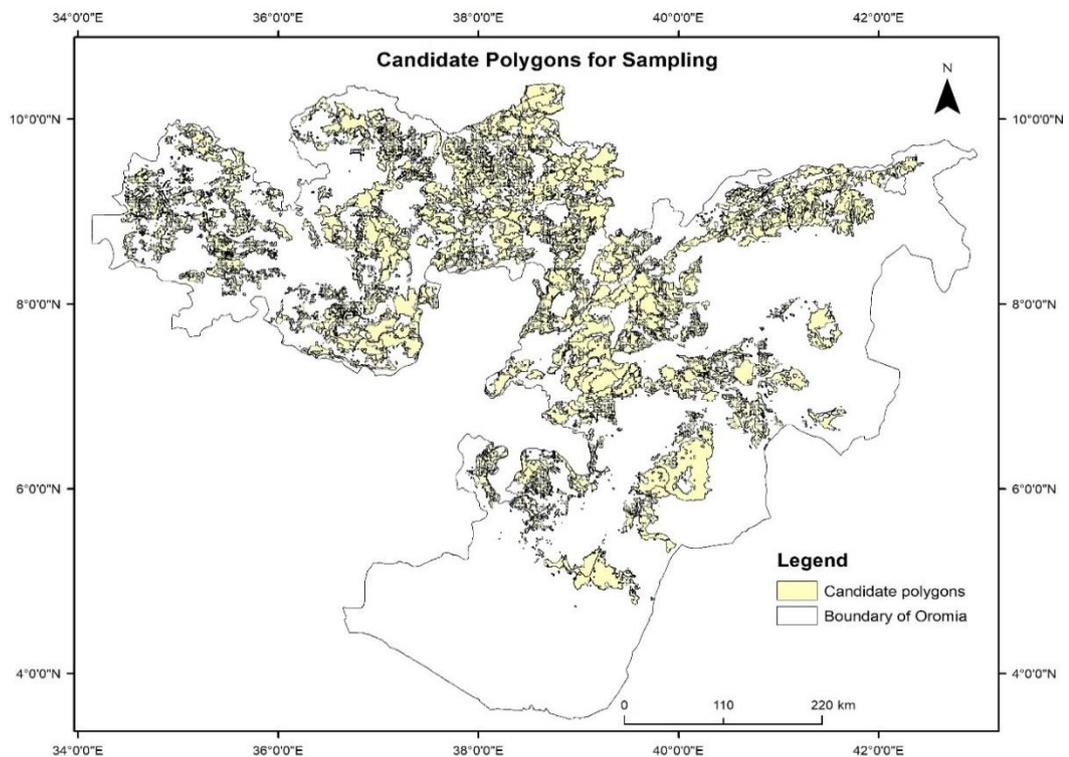


Figure 9: The four hundred candidate polygons for sampling

**4.1.3. Fifty Randomly Selected Polygons**

After intersecting the three maps and filtering the 400 candidate polygons, the primary sampling polygons were selected. From the filtered 400 candidate polygons, 50 polygons were randomly selected as a PSU, including 46 woredas from the administrative unit map, 12 zones from the CPS zone, and 22 from the livelihood zone as a sampling unit. Figure 6 above shows the spatial distribution of the 50 selected polygons and the 400 tiles in the study area, polygons denoted by black colour and sample tiles denoted by small red boxes.

**4.1.4. Sample tiles(400x) Eight in each polygon**

From the randomly selected polygons 400 tiles (1km x 1km), eight tiles from each polygon were randomly selected as SSU. Figure 6 above shows the spatial distribution of randomly selected polygons and the 400 tiles within polygons. Table 1 shows the total and the selected number of woredas, CPS zones, and LH zones as a sampling unit and the total number of polygons and tiles.

Table 1: The total and selected number of woredas and zones(strata) of stratification maps and total number of selected tiles.

Stratification maps	Total number of strata(zones/woredas)	Randomly selected woredas/zones	Candidate polygons after the intersection	Total selected sample polygons	Randomly selected tiles
CPS zones map	12	12	400	50	400 (8 tiles from each polygon)
LH zones map	60	22			
Administrative unit(woreda) map	249	46			

Table 2 shows the name of the 50 selected polygons in each strata map with their total selected counts. For example, the first row highlighted in red, shows that the number of strata according to the stratification maps that are found within one polygon, which includes one LH zone (ABW), six CPS zone (20, 33, 34, 35, 43, 57) and ten woredas (Amigna, Arsi Negele, Bale Gasera, Digeluna Tijo, Dodola, Ginnir Limuna Bilbilo, Merti, Robe and Tiyo).

Table 2: Counts of strata of the three strata maps in the selected polygons, column labels indicate CPS zone, the row labels indicate the LH zone code (Bold) and woredas.

Count of OBJI Column Labels	19	20	29	30	31	33	34	35	43	48	54	57	Grand Total
<b>ABW</b>		1				5	1	1	1			1	10
Amigna						1							1
Arsi Negele		1											1
Bale Gasera								1					1
Digeluna Tijo						1							1
Dodola												1	1
GINNIR									1				1
Limuna Bilbilo						1							1
Merti							1						1
Robe						1							1
Tiyo						1							1
<b>ACH</b>										1			1
Anfilo										1			1
<b>AMT</b>		1				1							2
Chelia		1											1
Wuchale						1							1
<b>AWT</b>		2	2										4
Ameya		1											1
Dendi			1										1
Weliso		1											1
Welmera			1										1
<b>BAT</b>			2		1								3
<b>BGE</b>											2	2	4
<b>BPA</b>									1				1
<b>CGC</b>		1	1										2
<b>CIE</b>			1										1
<b>DLS</b>												1	1
<b>DSM</b>		1											1
<b>HSF</b>		1											1
<b>MAS</b>		1	1			1							3
<b>NAP</b>				1									1
<b>NMT</b>		1											1
<b>OAP</b>									1		1		2
<b>RCS</b>		1											1
<b>RVM</b>				1	1		1						3
<b>SAW</b>		1				1							2
Adea Berga		1											1
Mulo						1							1
<b>SMC</b>		2					2						4
Gole Oda		1											1
Guba Qoricha		1					1						2
Meta							1						1
<b>WDH</b>		1											1
Horo		1											1
<b>WMS</b>												1	1
Seyo Nole												1	1
<b>Grand Total</b>	<b>7</b>	<b>10</b>	<b>4</b>	<b>2</b>	<b>2</b>	<b>8</b>	<b>4</b>	<b>1</b>	<b>3</b>	<b>1</b>	<b>3</b>	<b>5</b>	<b>50</b>

## 4.2. NDVI data Computation from Planet-Imagery

### 4.2.1. Time Series NDVI by Tiles (12 Months)

The time series NDVI images of the tiles were obtained after cloud masking and NDVI of individual months were performed. Then NDVI images of 2021 of the tiles were stacked and finally all tiles (400) NDVI time series image was exported from GEE platform for further analysis. Figure 10 shows an example of the stacked image of a tile.

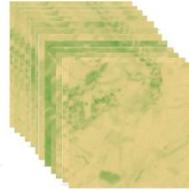


Figure 10: The time series NDVI image of a single tile of the study area (2021).

After exporting the 400 NDVI images, all images were organized using artificial georeferencing to make them all together as one file for simplifying the analysis. Figure 11 shows an example of the organized time series NDVI image of the 400 tiles (rows are sampled polygons (50) and columns are repeats of tiles (400)).

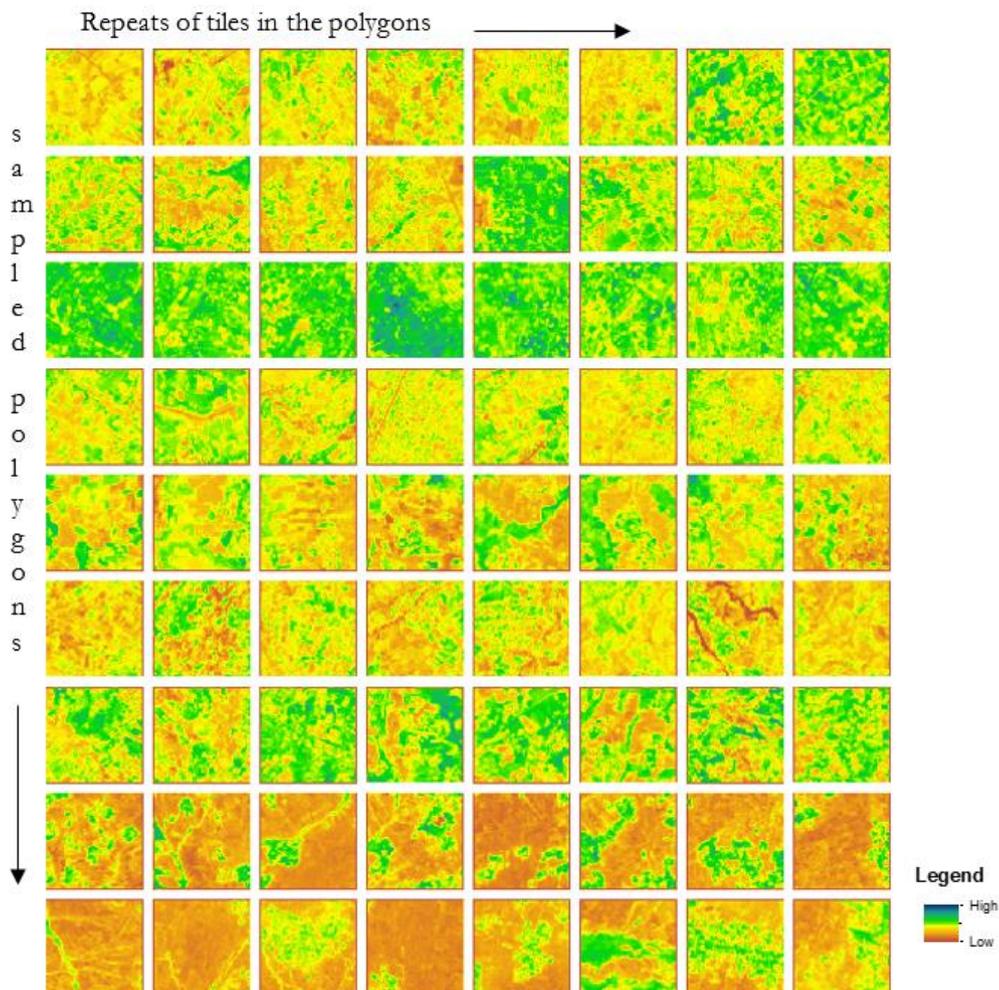


Figure 11: Examples of the organized (all in one file) NDVI image of the tiles rows indicates the sampled polygon and columns are repeats of tiles in the.

### 4.3. Time Series NDVI Data clustering

#### 4.3.1. Classified Tile Maps with 25 pixel-level clusters

After smoothing of the mosaiced NDVI series, a stack of the whole months of 2021 was clustered into 25 clusters using the ISODATA classification algorithm. The selected cluster and its signature NDVI profiles were used to analyse the land cover to identify areas of crop fields and other land covers. Figure 12 shows the 25 NDVI cluster classes of the tiles.

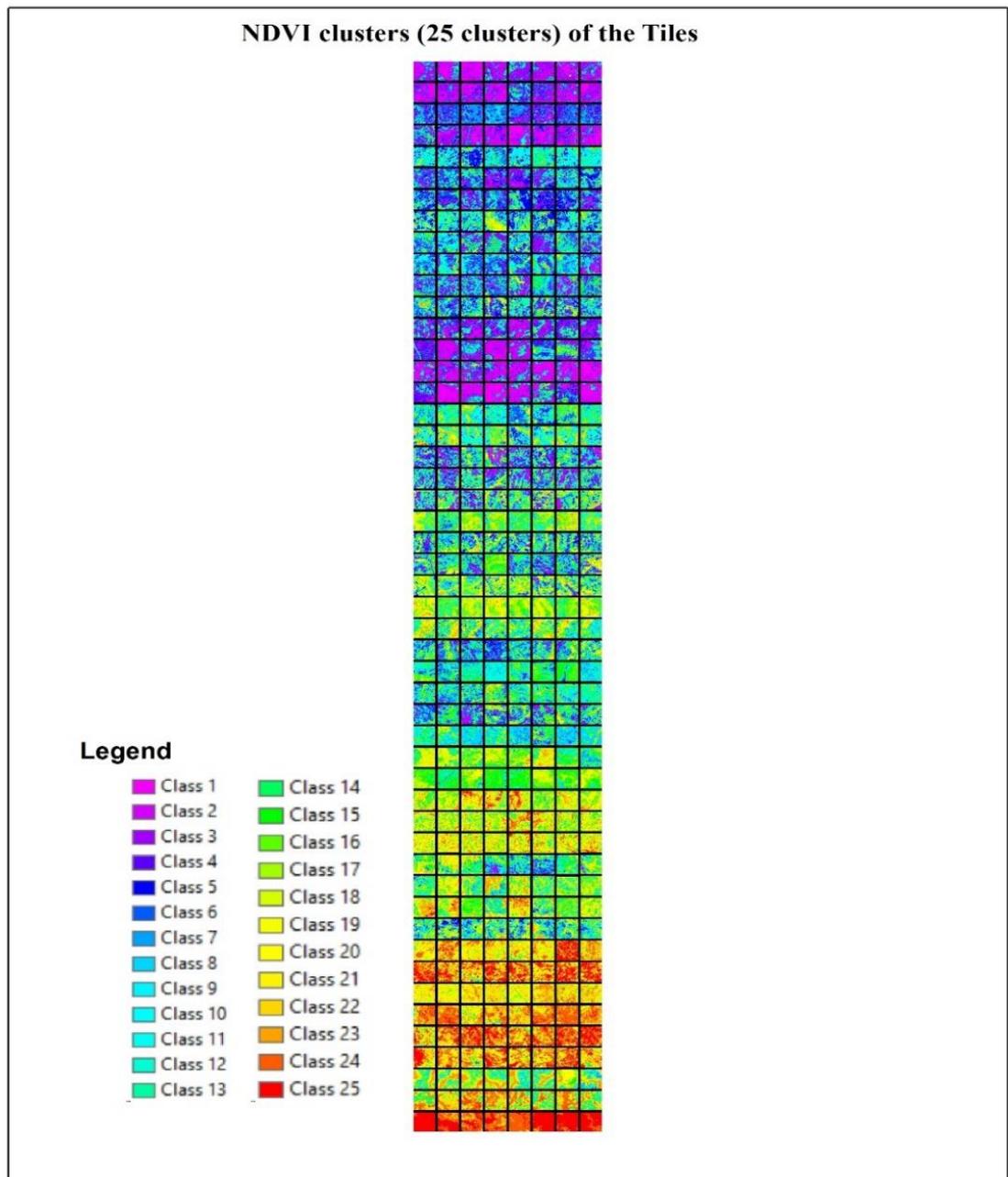


Figure 12: The 25 NDVI clusters of the tiles.

**4.3.2. Legends of the 25 NDVI Profiles by mapped cluster**

High-resolution imagery NDVI series can be used to determine the land cover types in smallholder areas (Begue et al., 2014). Figure 13 shows the legends of NDVI cluster profiles grouped based on the similarity of temporal characteristics of the NDVI. The groups of NDVI clusters were analysed using the legends of the NDVI profile to classify the land cover types.

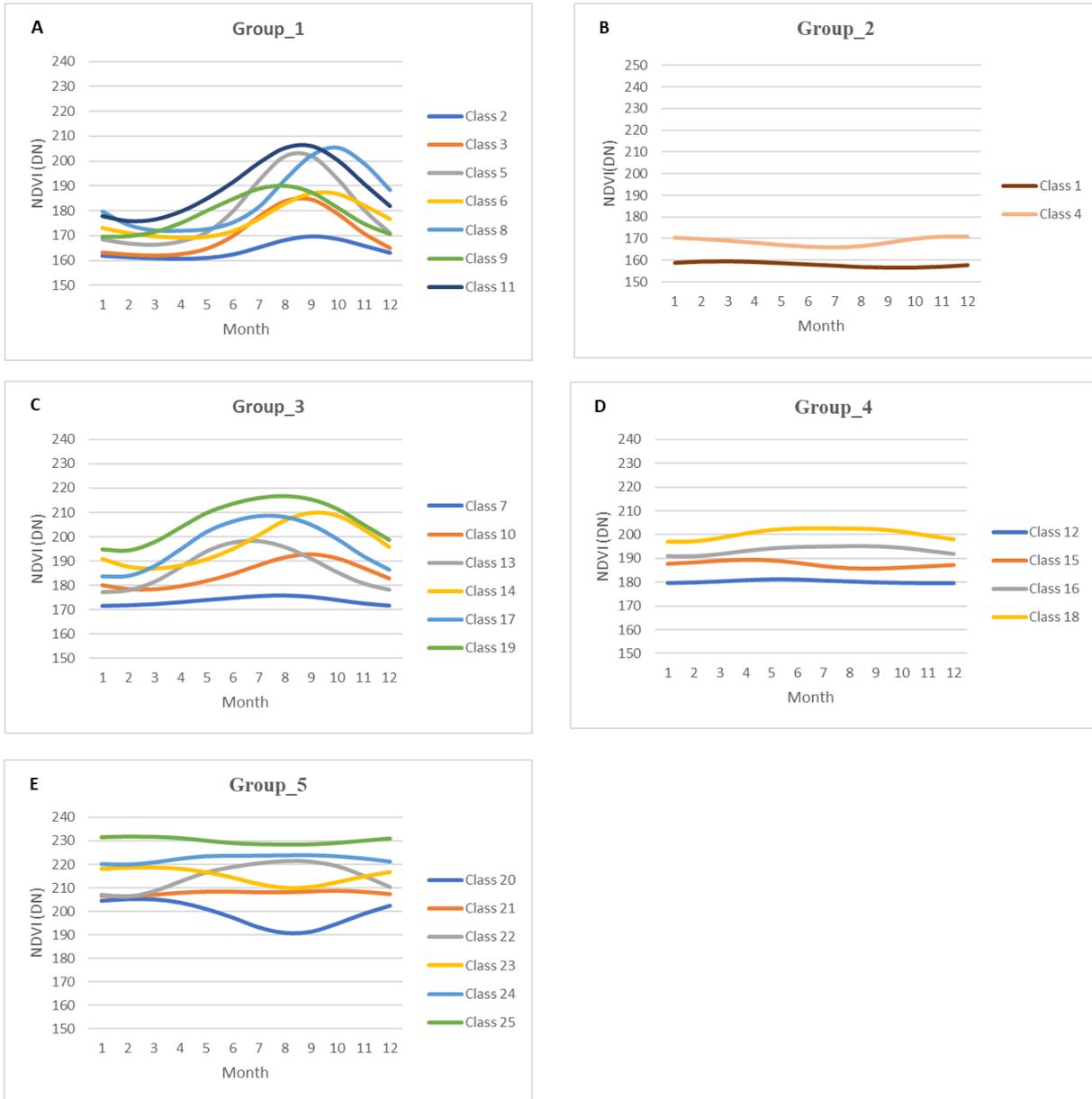


Figure 13: The 25 NDVI clusters grouped based on their similar characteristics, profiles (lines) denote the clusters, titles named groups denote the cluster after inspecting the GE image

A Google Earth (GE) imagery was used to inspect the representation of the profile by displaying each cluster and its image to identify the represented features of the clusters. Since the cluster represents groups of features with similar spectral characteristics, a single cluster may not indicate the same features on the GE imagery across the study area. Thus, the features represented by the NDVI profiles on the GE imagery were described in three ways (mostly, sometimes, and almost non times, see Annex-2) to identify the represented features. Mostly represent features, and the profile analysis was finally used to label the clusters. For example, Figure 13-A presents the temporal profile of crop fields because the NDVI profiles show a specific growing season, and Figure 14 indicates its representation on GE imagery that represents mostly crop fields. However, in Figure 13-B, the temporal profile indicates bare lands or water dominate areas because it indicates stable with low NDVI value throughout the year. Figure 15 shows its representation on GE indicates mostly bare lands/water body and sometimes fields without crops.

Based on the above NDVI cluster analysis, agricultural fields (crop fields) were identified. Figure 13-A shows seven different clusters, indicating the presence of crop field with one growing season; however, all clusters might not be within one growing season. Mohammed et al., 2020 indicated that the Oromia region (study area) is characterized by two growing seasons, Belg and Mehir. In some parts of the region, there is only one growing season characterized by dry and low-lying areas starting from July and ending in October (Mehir). Some parts of the region support the two growing seasons, characterized by wet and higher areas. The first season with low NDVI starts in April (Belg), and the second season which is longer than the first growing season, begins at the beginning of July (Mehir)(Mohammed et al., 2020). Therefore, the long growing season might be because of the data used was a monthly composite of one-year data that might generalize the lower NDVI value of the first growing season (Belg) while the data cleaning was performed.

The identified crop fields in Figure 13-A indicates growing season, which starts and ends in different months, covering both the Belg and Mehir seasons (Alemayehu et al., 2012). The growing season of cluster 6 and 8 starts in July and ends in October, which is the Mehir season known for producing maize and Sorghum. However, the growing season of cluster 2, 3, 5, 9, and 11 cover the growing range of both seasons that are known for supporting Belg and Mehir crop production, such as Sweet potato and Yams (Belg crops) and Teff and other grains, cereals and pulses(Mehir crops) (Mohammed et al., 2020).

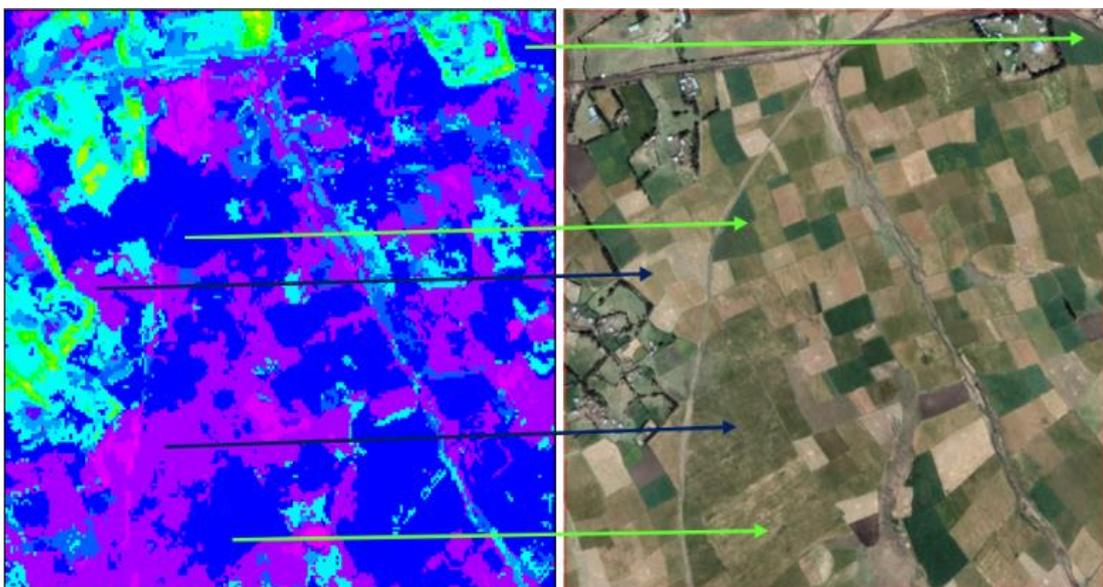


Figure 14: NDVI Clusters class and its represented features on GE image; Both the green and blue arrow indicate crop fields.

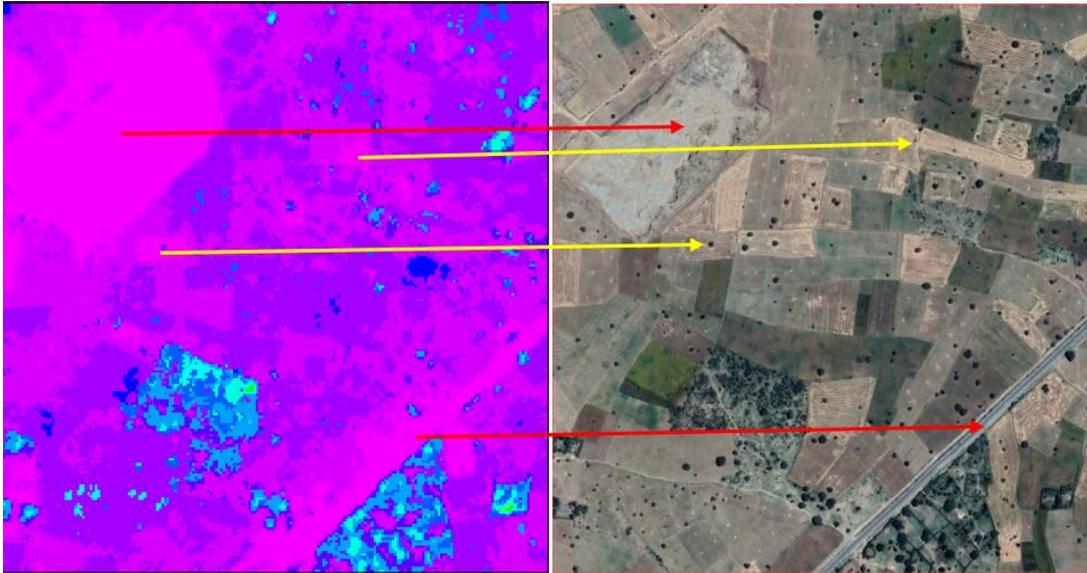


Figure 15: NDVI cluster and its represented features on GE image; bare land indicated by red arrow and yellow arrow indicates fields without crop (uncultivated land).

Other non-agricultural clusters were also analysed using the same procedure. The temporal behaviour of the clusters in Figure 13-C can be described as medium (i.e., greenness is not low as bare land) with increasing greenness in the wet seasons of the year, which indicates grassland areas. On the other hand, the temporal characteristics of the clusters in Figure 13-D and E are characterized to be stable and having a high greenness value throughout the year. Clusters in Figure 13-D indicate shrubs and trees with stable and relatively high greenness throughout the year, and clusters in Figure 13-E indicate dense forests with stable and high greenness throughout the year.

#### 4.4. Analyzing the Relationship Between Counts of Tiles by Tile Group and the Strata Map

##### 4.4.1. Area fraction of Clusters by Tiles

After annotating or labelling the NDVI clusters, the area fractions of the NDVI clusters were generated by converting the NDVI clusters into polygons (area fractions), which provided the area of the clusters. Table 3 presents some of the arranged data of the area fraction of the clusters by tiles according to the strata (zones and wordas) of strata maps. Data arrangement in this context means identifying zones and wordas of the tiles of NDVI clusters in the strata maps based on the attributes of the strata map.

Table 3: Examples of the area fraction of the clusters by tiles according to the strata maps (strata maps versus NDVI clusters, representing area fraction).

Strata Map			Area fraction of clusters in percent												
CPS	LH	Woredas	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	...
Zone	Zone														
31	24	Adea	7	37	27	5	0	5	8	0	1	3	0	5	...
31	24	Adea	22	38	5	8	0	1	11	0	4	2	0	6	...
31	24	Adea	33	55	9	2	0	1	1	0	0	0	0	0	...
31	24	Adea	7	32	31	5	2	2	8	0	7	2	0	3	...
31	24	Adea	1	8	55	2	3	4	7	0	11	2	0	4	...
31	24	Adea	6	29	47	3	1	2	5	0	3	2	0	3	...
31	24	Adea	24	27	33	5	0	1	6	0	1	0	0	2	...
31	24	Adea	9	27	31	7	0	2	13	0	3	1	0	5	...
31	136	Bora	23	39	27	4	1	2	4	0	0	0	0	1	...
31	136	Bora	10	37	35	3	1	8	2	0	1	1	0	1	...
31	136	Bora	12	37	31	8	3	2	6	0	0	0	0	0	...
31	136	Bora	21	52	10	8	0	1	8	0	0	0	0	1	...
31	136	Bora	3	4	15	16	2	17	10	1	9	5	0	9	...
31	136	Bora	2	17	50	2	5	14	4	0	2	3	0	2	...
31	136	Bora	15	47	11	11	0	2	9	0	2	1	0	1	...
31	136	Bora	18	37	35	3	1	2	4	0	0	0	0	1	...
30	113	Meiso	0	6	1	37	0	1	38	0	1	2	0	11	...
30	113	Meiso	0	3	7	34	0	17	32	0	3	0	0	4	...
30	113	Meiso	0	0	2	26	0	7	54	0	3	0	0	7	...
30	113	Meiso	0	1	6	3	2	18	42	0	27	0	0	1	...
30	113	Meiso	6	7	1	37	0	3	42	0	1	0	0	3	...
30	113	Meiso	3	20	8	37	0	9	17	0	2	2	0	3	...
30	113	Meiso	1	2	2	45	0	12	37	0	2	0	0	1	...
30	113	Meiso	1	1	1	47	0	4	40	0	0	0	0	6	...
...	.....	.....	..	..	..	...	..	..	...	..	..	..	..	..	...

Table 3 presents the area fraction of the clusters in each tile and in the strata (zones and wordas), of the strata maps. In each stratum, there is a repeat of tiles with minimum of eight tiles per strata. The strata of the tiles are listed in the strata map columns and the area fractions of the clusters in each tile (C1, C2, C3, C4, ...C25) are also listed under the columns of area fraction in percent, representing the clusters area. The cluster name was also labelled as it was analysed using both the NDVI profile and GE imagery in section 4.3.2

As indicated in Figure 13, the NDVI clusters were grouped into five groups based on their spectral similarity, and the annotation for each group was also given based on the NDVI profile analysis and GE imagery inspection. Therefore, the area fraction of the clusters can also be summarized based on the five

groups, including their area fraction according to the strata maps. Annex-3 shows the mean area fraction of the spectral similarity groups according to the three strata maps.

**4.4.2. Table of Grouped tiles (10) versus Clusters (25)**

Before computing the variability of the clusters among the strata maps, the relationships between the area fraction of the clusters (land cover) and strata maps were analysed. To prepare the data for analysing relationships between clusters and the strata maps, a hierarchical grouping of the tiles was performed using the cluster area fractions. The area fractions of the tiles were grouped into ten groups by tile using SPSS software. Table 4 shows the grouped tile versus clusters representing the average area fraction of grouped clusters. The clusters that show the crop field are highlighted in green.

Table 4: Grouped tiles (10) versus clusters (25) represents average area fraction in percent, the area fractions with crops filed highlighted in green.

Mean Percentage area of clusters in grouped tiles											
Clusters	1	2	3	4	5	6	7	8	9	10	Total
C1	23.86	2.29	0.65	4.95	0	0	0	0	0	0	2.88
C2	38.37	11.29	1.65	10.52	0	0	0	0	0	0	5.35
C3	16.06	44	3.23	2.52	0	0	0	0	0	0	4.75
C4	6.49	2.57	2.72	34.71	0	0	1	0	0	0	4.5
C5	0.97	15	3.56	0.24	3	0	0	0.5	0	0	3.06
C6	2.23	6.71	4.14	7.57	1	0	0	8.25	0	0	3.92
C7	5.31	5	5.18	20.76	0	20	3.4	1	0	0	5.75
C8	0	0.14	1.75	0.19	1	0	0	56.25	0	0	1.9
C9	2.23	6.43	4.25	3.9	1	1.67	0	2.25	0	0	3.76
C10	1.09	2.14	6.16	1.57	5	8	0.6	7.5	0	0	5.02
C11	0.06	0.57	3.94	0.71	55	0	0	2.75	0	0	3.2
C12	2.26	2.43	5.8	7.05	0	60	22.4	2.5	0	0	5.75
C13	0.46	0.29	3.85	1.9	2	1.67	1.8	4.5	0	0	3.14
C14	0	0	3.14	0.43	12	0	0	8.25	0	0	2.51
C15	0.69	0.86	5.51	1.9	0	7.33	50.6	1.5	0	0	5.04
C16	0.26	0.43	8.33	0.95	3	2.33	16.8	2	0	0	6.63
C17	0	0	3.54	0.29	4	0	0	1.25	0.17	0	2.72
C18	0	0	7.3	0.24	4	0	1.2	0.5	0.17	0	5.56
C19	0	0	4.31	0	1	0	0	0.75	3.56	0	3.43
C20	0	0	2.68	0.1	0	0	2	0	0	0	2.05
C21	0	0	6.31	0	5	0	0.2	0.5	1.28	0	4.85
C22	0	0	4.59	0	0	0	0	0.25	15.83	0.25	4.18
C23	0	0	3.63	0	2	0	0	0	3.56	0.25	2.91
C24	0	0	3.53	0	0	0	0	0	48.89	11.5	4.98
C25	0	0	0.77	0	0	0	0	0	27.11	88.5	2.69

**4.4.3. Crosstabulation Tables of the Strata Map**

As indicated above, the area fractions of the tiles were grouped into ten hierarchical groups, and crosstabulation tables were generated for each strata map. The crosstabulation tables show the counts of tiles in the group of tiles by strata. The result of crosstabulation of the three strata maps is presented as follows.

- **CPS zone crosstabulation table**

Table 5: Crosstabulation table of grouped tiles (10) versus Strata of the CPS zone map representing counts of tiles.

		Count; Average Linkage (Between Groups)										
CPS zone		1	2	3	4	5	6	7	8	9	10	Total
	19	0	0	56	0	0	0	0	0	0	0	56
	20	0	0	78	1	1	0	0	0	0	0	80
	29	16	0	5	11	0	0	0	0	0	0	32
	30	6	1	0	9	0	0	0	0	0	0	16
	31	13	2	1	0	0	0	0	0	0	0	16
	33	0	4	60	0	0	0	0	0	0	0	64
	34	0	0	28	0	0	3	1	0	0	0	32
	35	0	0	8	0	0	0	0	0	0	0	8
	43	0	0	16	0	0	0	4	4	0	0	24
	48	0	0	0	0	0	0	0	0	4	4	8
	54	0	0	19	0	0	0	0	0	6	0	25
	57	0	0	31	0	0	0	0	0	8	0	39
<b>Total</b>		<b>35</b>	<b>7</b>	<b>302</b>	<b>21</b>	<b>1</b>	<b>3</b>	<b>5</b>	<b>4</b>	<b>18</b>	<b>4</b>	<b>400</b>

Table 5 provides the ten groups of area fraction clusters by tiles among CPS zone, indicating the counts of the tiles in the tiles group versus strata of the CPS zone map.

- **LH zone Crosstabulation**

Table 6: Crosstabulation table of grouped tiles (10) versus Strata of the LH zone map representing counts of tiles.

		Counts of Average Linkage (Between Groups)										
		1	2	3	4	5	6	7	8	9	10	Total
LH Zone	ABW	0	4	69	0	1	0	0	4	2	0	80
	ACH	0	0	0	0	0	0	0	0	4	4	8
	AMT	0	0	16	0	0	0	0	0	0	0	16
	AWT	4	0	19	9	0	0	0	0	0	0	32
	BAT	19	1	2	2	0	0	0	0	0	0	24
	BGE	0	0	20	0	0	0	0	0	12	0	32
	BPA	0	0	7	0	0	0	1	0	0	0	8
	CGC	0	0	16	0	0	0	0	0	0	0	16
	CIE	0	0	8	0	0	0	0	0	0	0	8
	DLS	0	0	8	0	0	0	0	0	0	0	8
	DSM	0	0	8	0	0	0	0	0	0	0	8
	HSF	0	0	8	0	0	0	0	0	0	0	8
	MAS	0	0	23	1	0	0	0	0	0	0	24
	NAP	0	0	0	8	0	0	0	0	0	0	8
	NMT	0	0	8	0	0	0	0	0	0	0	8
	OAP	0	0	13	0	0	0	3	0	0	0	16
	RCS	0	0	8	0	0	0	0	0	0	0	8
	RVM	12	2	5	1	0	3	1	0	0	0	24
	SAW	0	0	16	0	0	0	0	0	0	0	16
	SMC	0	0	32	0	0	0	0	0	0	0	32
WDH	0	0	8	0	0	0	0	0	0	0	8	
WMS	0	0	8	0	0	0	0	0	0	0	8	
<b>Total</b>		35	7	302	21	1	3	5	4	18	4	400

Table 6 presents the ten groups of area fraction of clusters by tile among LH zones which indicates the counts of tile in the tile group versus strata of the LH zones map. From the ten hierarchical groups, most area fraction clusters by tiles are grouped under group 3.

• **Woredas Cross tabulation**

The result of the hierarchical clustering of tiles using clusters area fractions by tiles among Woredas is presented in Table 7.

Table 7: Crosstabulation table of grouped tiles (10) versus strata of the woreda map representing counts of tiles.

Counts of Average Linkage (Between Groups)												
Woredas		1	2	3	4	5	6	7	8	9	10	Total
	Abaya	0	0	7	0	0	0	0	0	0	1	0
Adea	7	1	0	0	0	0	0	0	0	0	0	8
Adea Berga	0	0	8	0	0	0	0	0	0	0	0	8
Ambo	0	0	16	0	0	0	0	0	0	0	0	16
Ameya	0	0	8	0	0	0	0	0	0	0	0	8
Amigna	0	0	8	0	0	0	0	0	0	0	0	8
Anfilo	0	0	0	0	0	0	0	0	0	4	4	8
Arsi Negele	0	0	7	0	1	0	0	0	0	0	0	8
Ayira	0	0	8	0	0	0	0	0	0	0	0	8
Bale Gasera	0	0	8	0	0	0	0	0	0	0	0	8
Bora	6	1	1	0	0	0	0	0	0	0	0	8
Boset	0	0	4	0	0	3	1	0	0	0	0	8
Chelia	0	0	8	0	0	0	0	0	0	0	0	8
Dawo	7	0	1	0	0	0	0	0	0	0	0	8
Dendi	1	0	0	7	0	0	0	0	0	0	0	8
Diga	0	0	8	0	0	0	0	0	0	0	0	8
Digeluna Tijo	0	2	6	0	0	0	0	0	0	0	0	8
Dodola	0	0	6	0	0	0	0	0	0	2	0	8
Dugdadewa	0	0	8	0	0	0	0	0	0	0	0	8
Gemechis	0	0	8	0	0	0	0	0	0	0	0	8
GINNIR	0	0	4	0	0	0	0	4	0	0	0	8
Girawa	0	0	8	0	0	0	0	0	0	0	0	8
Gole Oda	0	0	8	0	0	0	0	0	0	0	0	8
Guba Qoricha	0	0	16	0	0	0	0	0	0	0	0	16
Horo	0	0	8	0	0	0	0	0	0	0	0	8
Ilu	5	0	1	2	0	0	0	0	0	0	0	8
Leqa Dulecha	0	0	8	0	0	0	0	0	0	0	0	8
Liben	0	0	5	0	0	0	3	0	0	0	0	8
Limuna Bilbilo	0	1	7	0	0	0	0	0	0	0	0	8
Meda We Labu	0	0	7	0	0	0	1	0	0	0	0	8
Meiso	0	0	0	8	0	0	0	0	0	0	0	8
Mertie	0	0	8	0	0	0	0	0	0	0	0	8
Meta	0	0	8	0	0	0	0	0	0	0	0	8
Metarobi	0	0	15	1	0	0	0	0	0	0	0	16
Mulo	0	0	8	0	0	0	0	0	0	0	0	8
Odo Shakiso	0	0	6	0	0	0	0	0	2	0	0	8
Omonada	0	0	8	0	0	0	0	0	0	0	0	8
Robe	0	0	8	0	0	0	0	0	0	0	0	8
Seyo Nole	0	0	8	0	0	0	0	0	0	0	0	8
Sude	0	0	8	0	0	0	0	0	0	0	0	8
Tiyo	0	1	7	0	0	0	0	0	0	0	0	8
Uraga	0	0	7	0	0	0	0	0	9	0	0	16
Weliso	0	0	8	0	0	0	0	0	0	0	0	8
Welmera	3	0	3	2	0	0	0	0	0	0	0	8
Wuchale	0	0	8	0	0	0	0	0	0	0	0	8
Zeway Dugda	6	1	0	1	0	0	0	0	0	0	0	8
Total		35	7	302	21	1	3	5	4	18	4	400

**4.4.4. Pearson Chi-Square (X<sup>2</sup>) Result**

The relationship between the counts of tiles and the strata maps was examined for each of the maps through the Pearson chi-square test. The area fractions of the clusters and strata of the strata maps are represented by a crosstabulation table (row x column) in the above section as a categorical variable. Thus, Pearson’s chi-square test statistics can be applied to test the relationships between the row and column variables(Shih and Fay, 2017). The results of the Pearson chi-square tests of the strata maps are presented in Table 8.

Table 8: The three individual Chi-Square tests result to examine if counts of tiles are significantly related to the stratification map.

Pearson Chi-Square Tests				
		Value	df	Asymptotic Significance
Pearson Chi-Square (X <sup>2</sup> )	CPS zone	876	99	<.001
	LH zone	907	189	<.001
	Woredas	1508	405	<.001

Table 8 shows the result of the X<sup>2</sup> test of CPS zones, LH zones, and woredas map. Value X<sup>2</sup> = 876 and asymptotic significance P<0.001 shows there is a significant relationship between counts of tiles and the CPS zone map. Value X<sup>2</sup>= 907, and significance P<0.001 indicates a significant relationship between counts of tiles and the LH zone map. Value X<sup>2</sup>=1508 and significance level P<0.001 also indicate a significant relationship between counts tiles and woredas map. The chi-square test shows that all three strata maps have a significant relationship with the counts of tiles (NDVI clusters).

**4.5. Analyzing the Variability of NDVI Clusters (Crop Field) by Strata Map**

**4.5.1. Analysis of Variance (ANOVA) by Strata Map**

The average area percentage of the area fractions of the clusters obtained from the NDVI clusters was statistically evaluated by ANOVA with Post-hoc Tukey-HSD and Hochberg’s GT2 test. The ANOVA analysis results were helpful in determining the NDVI cluster differences of strata according to the strata maps and identifying strata maps with maximum variability between strata and minimum variability within strata. Firstly, three one-way ANOVAs were performed according to the strata maps and area fraction of the clusters to identify the significant differences for each strata map. The ANOVA results of the strata maps determined a statistically significant difference between the area fraction of clusters of strata of the strata maps with a p-value less than 0.05 (p<0.05).

The ANOVA test only determines whether there is a significant difference between the area fraction of clusters according to the strata of the maps. However, it is important to determine which strata map pairs of strata mostly show differences, which helps to indicate strata maps with higher heterogeneity between strata (zones and woredas). Therefore, secondly, a Tukey HSD and Hochberg’s GT2 post hoc tests were included for multiple comparisons of the area fraction of clusters values according to the strata of the strata maps. As a result, for the CPS zone and LH zone maps, a Hochberg GT2 test was used for multiple comparisons since the group sizes are not equal. In contrast, the Tukey HSD test was used for multiple comparisons of woredas because group sizes are almost equal for all woredas.

According to the statistical relationship between the strata and the area fraction of clusters, strata maps with better heterogeneity between strata were identified among the statistically significant strata. The result of Hochberg's GT2 test shows that, of all the pairs of strata significance differences of the NDVI clusters, 55 percent of the CPS zone pairs of strata had statistically significant differences, and 35% of the LH zone pairs of strata showed a significant difference at a 5% significance level. Moreover, the Tukey HSD test (Annex 4-C) shows that, of all the pairs of strata significance difference, 39% of the woreda pairs indicated a significant difference at a 5% significance level.

The significant differences of pairs of strata of the area fraction of the crop fields cluster were checked to determine the significant differences of the crop fields cluster according to the strata maps. The result shows that of all pairs of strata significance difference of crop fields cluster (Annex 4-A and B), 60% of the CPS zone pairs of strata had a significant difference, 38% of the LH zone pairs of strata had a significant difference, and 51% of the woreda pairs of strata had a significant difference at 5 % significance level (Annex 4-A and B).

For this study, it is also important to determine which strata map shows the minimum within-groups and maximum between-groups differences in capturing the crop fields, which helps to indicate the homogeneity of the strata within groups and the heterogeneity of strata in representing crop fields. Therefore, thirdly, an ANOVA was checked with one variable using only the area fraction of clusters that were grouped as crop fields group. This process provides a general ANOVA result that shows crop field differences within groups and between groups. Table 9 shows the ANOVA results of crop fields according to the three strata maps.

Table 9: The within and between groups (strata) difference in the crop field clusters of the three strata maps.

ANOVA						
Crop field clusters differences						
Strata Map		Sum of Squares	df	Mean Square	F	Sig.
	Between Groups	136891.44	11	12444.676	42.946	<.001
<b>CPS zone</b>	Within Groups	112433.457	388	289.777		
	Total	249324.897	399			
	Between Groups	141246.291	21	6726.014	23.524	<.001
<b>LH zone</b>	Within Groups	108078.606	378	285.922		
	Total	249324.898	399			
	Between Groups	198351.585	45	4407.813	30.611	<.001
<b>Woredas</b>	Within Groups	50973.313	354	143.992		
	Total	249324.898	399			

## 5. DISCUSSION

### 5.1. On Sampling Scheme

Sampling a very high-resolution satellite image is helpful in estimating the land cover area and accuracy of landcover maps, especially when an area's coverage is extensive (Gallego, 2011). In this research, the satellite image was sampled to compare three stratification maps based on the NDVI clusters (land cover) variability, particularly crop field clusters variability between strata. The strata maps were intersected to filter the candidate polygons with an area larger than 100 km<sup>2</sup>, which were used in sampling. Four hundred polygons identified as a candidate and a two-stage sampling scheme was employed for representing land cover since a two-stage sampling scheme gives an alternative to complete stratification if the study area is too broad (Gallego, 2004; Wang et al., 2018). The sampling scheme stages used in this research were similar to the ones followed by Gallego (2011); however, the process was different. For example, they used a grid of certain kilometres for the selection of PSU for the first stage sampling, whereas the PSU selected in this research was based on random polygons.

The sample polygons were chosen in the first stage as PSU, and the 1km-by-1km tiles were randomly selected as SSU in the second stage. The reason of using a 1km-by-1km tiles was to minimize the effects of cloud and consider the smallest unit of measurement of the CPS zone. 400 tiles (SSU) were selected to sample the study area from randomly selected 50 polygons (PSU), which allows us to generate polygons or strata with unique values of the CPS zone, LH zone, and woreda map. The sample size (i.e., tiles) in the strata of the strata map was not equal because the strata maps had different method of classification with different parameters that resulted in different area sizes and strata. As indicated in Table 1 the result showed that the number of sampled strata from the total number of strata for the CPS zone, LH zone, and woreda map were 12, 22, and 46 strata, respectively. Therefore, the sampling size for the CPS zone and LH zone are not equal since the area sizes of the strata were large and covers wider areas, which indicates more generalization of the strata map, whereas the sampling size for woredas is almost equal because of the area size of the woreda was small, which indicates the woreda map is less generalized and provides samples from a small area.

### 5.2. On NDVI Data Computation

Time series NDVI is one of the most well-known methods for crop monitoring and spatial distribution of crops (Veloso et al., 2017). In this study, NDVI was used as a reliable land cover indicator. The NDVI of the high-resolution imagery is significant for identifying crop areas in smallholder areas since small-scale farming is more difficult to detect with coarse resolution imagery (Begue et al., 2014). As it was proven that the high-resolution imagery from landsat-8 and Sentinel-2 being the main source of data to capture crop areas in recent years (Werner et al., 2019). However, the spatial resolution (4.77m) of Planet imagery as twice as the spatial resolution of the sentinel 2 (10 m), NDVI from Planet could identified crop fields of the study areas. The NDVI images were obtained from the planet NICFI tropical mosaics through the GEE platform. Even though the released image was a monthly composite, there was cloud cover in some months in 2021, so cloud masking was performed individually for each month since the cloud cover was different. Here we can see the advantages of planet revisit time (daily) in obtaining some cloud free images in rainy seasons of the study area. The NDVI was also computed first for each individual month, and all months (2021) of the tiles were stacked and exported from the GEE platform. The result showed that a single NDVI based technique could be able to accurately map crop fields in fragmented landscapes.

### 5.3. On NDVI Data clustering

An unsupervised classification method using the ISODATA algorithm was used to produce the land cover using a high-resolution NDVI of 12 months of 2021. The ISODATA algorithm uses clustering techniques to automatically cluster similar pixels into groups (Kvamme et al., 2019).

Using the ISODATA clustering algorithm, 25 NDVI clusters were produced. The result indicated that the high-resolution Planet NDVI images could be used to identify or capture crop fields in fragmented areas. The different land cover types of the area can be distinguished by their profiles as described in section 4.3.2 and grouped into five land cover types. The NDVI clusters were identified by their feature representation using temporal profiles and GE imagery inspection as described in section 4.3.2. This is a simple method for optimizing the number of clusters, which in this study was five. De Bie et al. (2008), on the other hand, proposed using separability values (i.e., minimum, and average separability) to figure out the optimum number of purposeful ecological unit. For future research, separability analysis may be a better option for figuring out the ideal number of clusters. The result showed that the same (stable) NDVI profiles with low NDVI values indicated bare lands; the same (stable) NDVI temporal profiles with high NDVI values represented green vegetation (trees or shrubs land and dense forests); and the temporal profiles, which can be described as medium (i.e., the greenness is not as low as bare lands) with increasing profile lines in the wet season, indicated grassland areas. The stable temporal profiles with a specific growing season (high NDVI values) represented agricultural areas (crop fields), which indicates one growing season. However, the clusters might not be limited to just one growing season. As indicated by Mohammed et al. (2020), the study area (Oromia region) is characterized by a two-growing season, Belg and Mehir. In some parts of the region, there is only one growing season characterized by dry and low-lying areas, which starts in July and ends in October (Mehir), whereas some other parts of the region support two growing seasons characterized by wet and higher areas. The first growing season with low NDVI starts in April (Belg) and the second season, which is longer than the first season, begins at the beginning of July (Mehir) (Mohammed et al., 2020). As a result, the one growing season indicated by the profile, particularly the long growing season that began in April and ended in October, could be because the data used was a monthly composite of one-year data, which could generalize the NDVI value of the first growing season (Belg) and the upper enveloping might be overdone. For this research, identifying NDVI clusters was considered to classify the land cover to generate the area fractions, particularly for crop fields, because the final analysis was done based on the area fraction of the crop fields. However, future studies can identify crop areas in the two-growing season by improving the cloud masking process and the upper enveloping process.

### 5.4. On The Relationship Between Counts of Tiles by Tile Group and the Strata Maps

Chi-square test was used to test the relationships between the land cover (NDVI clusters), and the strata maps. Pearson chi-square test was widely used to test the relationship between categorical variables (Shih and Fay, 2017).

In this study, the area fraction of the NDVI clusters were arranged and grouped by tiles to analyse the relationship between the NDVI clusters (land cover) and strata maps for generalized strata map analysis. A hierarchical clustering approach was used to group the tiles using the area fractions of the clusters. Hierarchical clustering approaches try to identify relatively homogeneous groups of variables based on selected characteristics and detect clusters within a dataset to reduce clustering while maintaining the

structure of the data (Gonçalves et al., 2008). The 400 tiles were grouped into ten groups of tiles by tile group using area fractions of the clusters, which represents the average area fraction of the clusters in each group of tiles by tile group. These grouped tiles were generated according to the strata maps (CPS zone, LH zone, and woreda map) using crosstabulation, and provide a crosstabulation table that described strata versus grouped tiles by tile-group, representing counts of tiles. This means that the clusters area fraction by grouped tile and the strata of the strata maps were presented in row by column as categorical data. Pearson's chi-square test statistics can be applied to test the relationships between the rows and columns variables (Shih and Fay, 2017). As indicated in Table 9, the chi-square result showed there is a significant relationship between counts of tiles by tile group and strata maps at a 5% significance level, which rejects the null hypothesis ( $H_0-1$ ) that there was no significant relationship at 5% between counts of tiles by tile-group and strata maps. This indicates that the NDVI clusters or land cover are significantly related to the selected existing stratification method. However, the significant relationship between NDVI clusters and the three maps does not mean that the three strata maps are the same in determining land cover. And it could not be possible to identify which strata map is relatively not biased in capturing the land covers or crop areas.

### 5.5. On The Variability of Crop field Clusters by Strata Map

The last part of this study was analysing the variability of the area fraction of crop field NDVI clusters and strata maps. An ANOVA was used to analyse the variability of the area fraction of the clusters and strata maps. ANOVA measures the significance of the differences between groups (Cilek and Uslu, 2022; Jackson and Ferguson, 1972).

The statistical analysis result indicated a statistically significant difference between the strata in the NDVI clusters of the strata maps at a 5% significance level. However, for this study, it was important to determine which strata maps had a minimum difference within and a maximum difference between groups or strata. Further ANOVA analysis was done to determine the difference between pairs of strata. A statistical post hoc analysis was employed to identify differences between groups (Cilek and Uslu, 2022). The result indicated that of all significantly different pairs of strata of the strata maps, the CPS zone map had the highest significance differences (60%) in the area fraction of crop field clusters at a 5% significance value. This rejects the null hypothesis ( $H_0-2$ ), which was that the counts (percentage) of significantly different pairs of strata in crop field clusters at a 5% significance value is lower for the CPS zone map relative to the others. These pairs of strata significance differences indicated that the strata of the CPS zone map were relatively more heterogeneous than the others, which means neighbouring strata were better identified and could be unbiased in representing and better in capturing crop fields than the other strata maps. Mohammed et al. (2020), also describes, CPS zones consider the "variability of the cropping calendar by stratifying areas into the similar vegetation response that clustering the landscape on a pixel-by-pixel basis into homogeneous areas exhibiting similar temporal patterns in the NDVI." Previous studies achieved promising results in estimating the field fraction of crop areas using the CPS zone strata map. In the same study area in Ethiopia, they obtained an adjusted  $R^2$  of 91.4%, estimating field fraction in the reported crop areas in a study by Mohammed et al. (2020) and another study by Khan et al. (2010), in a different study area (southern Spain) and with specific three crop types, the authors achieved an adjusted  $R^2$  of 98.8%, 97.5%, and 76.5 %.

The woreda map had 51% of all significantly different pairs of strata in crop field cluster area fraction. The LH zone map had the lowest number of significantly different pairs of strata in area fraction of crop field

clusters, which was 38%. The LH zone map's least significant difference in the area fraction of crop field clusters indicated that the strata does not mean that less effective in representing samples or crop fields.

On the other hand, the result of the ANOVA, as indicated in Table 9 that was analysed without the post hoc test, indicated that the woreda map had the lowest (minimum) within-group differences than the others, and the CPS zone map had the highest (maximum) within-group differences. This indicated that the strata of the woreda map were homogeneous within strata, which could provide accurate samples. However, it might have many samples to be surveyed, which could take large amounts of money and time. That could be a result of many areas not being surveyed over a given time. The Ethiopian Central Statistics Agency (CSA) (CSA, 2021), agricultural sampling survey report on land utilization indicated that the total enumeration areas were not surveyed effectively. They planned 2768 enumeration areas for the country, but they achieved 2371 enumeration areas. This might be because of many numbers of the sampling areas and could affect the data quality.

## 6. CONCLUSION AND RECOMONDATION

This research compared three area stratification methods for improving the estimation of crop area statistics in fragmented agricultural landscapes based on the area fraction of the crop field clusters variability between strata. The three existing strata maps were intersected to select sample areas, which were used to select tiles of the sampled images to generate an area fraction of the NDVI clusters.

The area fractions of the clusters were obtained from high-resolution monthly composite imagery of one year (2021). The study identified that the CPS zone map had the highest variability between strata in crop field clusters, 60% of pairs of strata were significantly different. This indicates that the CPS zone map considers the spatial variability between strata that can identify uniform areas and help to generalize large areas, which could minimize the cost and time of surveying. On the other hand, the administrative unit map had minimum variability within strata than the others. This indicates that the administrative unit map provides samples from a small area that can accurately represent the area. However, the sample areas did not consider the spatial variability of the area, which might affect the representation of the samples over the wider area, and many sampling areas could take a large amount of money and time. The LH zone map had indicated relatively low variability between strata and relatively higher within strata variability. It does not mean that the LH zone map is less effective in capturing agricultural areas (crop fields). This might be because this study used only one indicator (i.e., NDVI) and/or sample area selection. In future studies, with additional indicators, can be examined.

Stratification with an administrative unit map might give the sample in a small area that can accurately represent the area. However, since the sample areas are administrative boundaries that may not consider the spatial variability of the area, which may result in the identification of areas without crop areas or dense crop areas that can affect the representation of the area. On the other hand, since CPS zones consider the spatial variability, it can identify relatively uniform areas and help to select stratification with the same characteristics. However, there might be some noise (miss clustered) areas with cropping areas that might show the same characteristics and affect the selected samples and its representations, especially in fragmented landscapes.

Therefore, the stratification is supposed to be representative, relatively uniform within, and generally different from the neighbouring strata to reduce the uncertainty of crop area estimation. Integrating both the CPS zone map and the administrative unit map could provide the proper stratification method. The findings support that integrating the CPS zone map and the administrative unit map can provide a stratification method for fragmented landscapes like Ethiopia, reducing the uncertainty in crop area estimation.

## LIST OF REFERENCES

- Adugna Aynalem, 2021. OROMIYA: Demography and Health [WWW Document]. URL [www.EthioDemographyAndHealth.Org](http://www.EthioDemographyAndHealth.Org) (accessed 12.9.21).
- Aguilar, R., Zurita-Milla, R., Izquierdo-Verdiguier, E., de By, R.A., 2018. A cloud-based multi-temporal ensemble classifier to map smallholder farming systems. *Remote Sens.* 10. <https://doi.org/10.3390/rs10050729>
- Alemayehu, S.T., Paul Dorosh, Sinafikeh Asrat, 2012. Crop Production in Ethiopia: Regional Patterns and Trends Summary of ESSP II Working Paper 16, Research Note 11. International Food Policy Reaserch Institute, Addis Ababa. <https://doi.org/https://reliefweb.int/sites/reliefweb.int/files/resources/essprn11.pdf>
- Ali, A., de Bie, C.A.J.M., Skidmore, A.K., Scarrott, R.G., Hamad, A., Venus, V., Lymberakis, P., 2013. Mapping land cover gradients through analysis of hyper-temporal NDVI imagery. *Int. J. Appl. Earth Obs. Geoinf.* 23, 301–312. <https://doi.org/10.1016/J.JAG.2012.10.001>
- Barrett, C.B., 2010. Measuring Food Insecurity. *Science* (80-. ). 327, 825–828. <https://doi.org/10.1126/science.1182768>
- Bauer, M.E., Hixson, M.M., Davis, B.J., Etheridge, J.B., 1978. Area Estimation of Crops by Digital Analysis of Landsat Data. *Photogramm. Eng. Remote Sensing* 44, 1033–1043. [https://doi.org/http://www.asprs.org/wp-content/uploads/pers/1978journal/aug/1978\\_aug\\_1033-1043.pdf](https://doi.org/http://www.asprs.org/wp-content/uploads/pers/1978journal/aug/1978_aug_1033-1043.pdf)
- Begue, A., Vintrou, E., Saad, A., Hiernaux, P., 2014. Differences between cropland and rangeland MODIS phenology (start-of-season) in Mali. *Int. J. Appl. Earth Obs. Geoinf.* 31, 167–170. <https://doi.org/10.1016/j.jag.2014.03.024>
- Central Statistics Agency(CSA), 2015. Ethiopia - Agricultural Sample Survey 2014-2015 (2007 E.C), Ethiopian Central Statistics Agency.
- Central Statistics Agency of Ethiopia, 2015. Ethiopia - Agricultural Sample Survey 2014-2015 (2007 E.C). Addis Ababa.
- Chen, X., Xu, Y., Yang, J., Wu, Z., Zhu, H., 2020. Remote sensing of urban thermal environments within local climate zones: A case study of two high-density subtropical Chinese cities. *Urban Clim.* 31. <https://doi.org/10.1016/J.UCLIM.2019.100568>
- Cilek, M.U., Uslu, C., 2022. Modeling the relationship between the geometric characteristics of urban green spaces and thermal comfort: The case of Adana city. *Sustain. Cities Soc. J. homepage.* <https://doi.org/10.1016/j.scs.2022.103748>
- Craig, M., Atkinson, D., 2013. A Literature Review of Crop Area Estimation, accessed on Nov 2021. [https://doi.org/https://www.fao.org/fileadmin/templates/ess/documents/meetings\\_and\\_workshops/GS\\_SAC\\_2013/Improving\\_methods\\_for\\_crops\\_estimates/Crop\\_Area\\_Estimation\\_Lit\\_review.pdf](https://doi.org/https://www.fao.org/fileadmin/templates/ess/documents/meetings_and_workshops/GS_SAC_2013/Improving_methods_for_crops_estimates/Crop_Area_Estimation_Lit_review.pdf)
- CSA, (Central Statistical Agency), 2021. Report on Land Utilization by Private Peasant Holdings in 2020/2021 Meher Cropping Season, Statistical Bulletin. Addis Ababa.
- Dahal, D., Wylie, B., Howard, D., 2018. Rapid crop cover mapping for the conterminous United States. *Sci. Rep.* 8, 1–12. <https://doi.org/10.1038/s41598-018-26284-w>
- De Bie, C.A., Khan, M.R., Toxopeus, A.G., Venus, V., Skidmore, A.K., 2008. Hypertemporal image analysis for crop mapping and change detection. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. - ISPRS Arch.* 37. [https://doi.org/https://www.researchgate.net/publication/313102904\\_Hypertemporal\\_image\\_analysis\\_for\\_crop\\_mapping\\_and\\_change\\_detection](https://doi.org/https://www.researchgate.net/publication/313102904_Hypertemporal_image_analysis_for_crop_mapping_and_change_detection)
- de Bie, C.A.J.M., Khan, M.R., Smakhtin, V.U., Venus, V., Weir, M.J.C., Smaling, E.M.A., 2011. Analysis of multi-temporal SPOT NDVI images for small-scale land-use mapping. *Int. J. Remote Sens.* 32, 6673–6693. <https://doi.org/10.1080/01431161.2010.512939>
- Eggen, M., Ozdogan, M., Zaitchik, B., Simane, B., 2016. Land Cover Classification in Complex and Fragmented Agricultural Landscapes of the Ethiopian Highlands. *Remote Sens.* 8, 1020.

- <https://doi.org/10.3390/rs8121020>
- ERDAS, 2003. ERDAS (2003) Erdas Field Guide. 7th Edition, Leica Geosystems, GIS and Mapping LLC, Atlanta. - References - Scientific Research Publishing [WWW Document]. URL [https://www.scirp.org/\(S\(i43dyn45teexjx455qlt3d2q\)\)/reference/ReferencesPapers.aspx?ReferenceID=1268829](https://www.scirp.org/(S(i43dyn45teexjx455qlt3d2q))/reference/ReferencesPapers.aspx?ReferenceID=1268829) (accessed 1.17.22).
- FAO, 2021. ESS: CHAPTER 7. PREPARATION OF THE FRAMES [WWW Document]. URL <https://www.fao.org/economic/the-statistics-division-ess/world-census-of-agriculture/conducting-of-agricultural-censuses-and-surveys/chapter-7-preparation-of-the-frames/en/> (accessed 12.26.21).
- FAO, 2018. The future of food and agriculture Alternative pathways to 2050. Rome <https://www.fao.org/3/I8429EN/i8429en.pdf>.
- FAO, 2017. MASTER SAMPLING FRAME(MSF) FOR AGRICULTURAL STATISTICS 13–14.
- FAO, 2015. HANDBOOK ON Master Sampling Frames for Agricultural Statistics Frame Development, Sample Design, and Estimation.
- FAO, WFP, 2007. S P E C I A L R E P O R T FAO/WFP CROP AND FOOD SUPPLY ASSESSMENT MISSION TO ETHIOPIA. Rome.
- FEWSNET, 2021. Livelihoods | Famine Early Warning Systems Network [WWW Document]. URL <https://fews.net/sectors-topics/sectors/livelihoods> (accessed 12.26.21).
- Fritz, S., See, L., Bayas, J.C.L., Waldner, F., Jacques, D., Becker-Reshef, I., Whitcraft, A., Baruth, B., Bonifacio, R., Crutchfield, J., Rembold, F., Rojas, O., Schucknecht, A., Van der Velde, M., Verdin, J., Wu, B., Yan, N., You, L., Gilliams, S., Mücher, S., Tetrault, R., Moorthy, I., McCallum, I., 2019. A comparison of global agricultural monitoring systems and current gaps. *Agric. Syst.* 168, 258–272. <https://doi.org/10.1016/j.agry.2018.05.010>
- Gallego, Francisco, J., 1999. CROP AREA ESTIMATION IN THE MARS PROJECT. <https://doi.org/https://www.semanticscholar.org/paper/CROP-AREA-ESTIMATION-IN-T>
- Gallego, F.J., 2015. Area Sampling Frames for Agricultural and Environmental Statistics third main title line. European union.
- Gallego, F.J., 2011. The efficiency of sampling very high resolution images for area estimation in the European Union. *Int. J. Remote Sens.* 33, 1868–1880. <https://doi.org/10.1080/01431161.2011.602993>
- Gallego, F.J., 2004. Remote sensing and land cover area estimation. *Int. J. Remote Sens.* 25, 3019–3047. <https://doi.org/10.1080/01431160310001619607>
- GHI, 2021. Ethiopia - Global Hunger Index (GHI) - peer-reviewed annual publication designed to comprehensively measure and track hunger at the global, regional, and country levels [WWW Document]. URL <https://www.globalhungerindex.org/ethiopia.html> (accessed 11.28.21).
- Gonçalves, M.L., Netto, M.L.A., Costa, J.A.F., Zullo Júnior, J., 2008. An unsupervised method of classifying remotely sensed images using Kohonen self-organizing maps and agglomerative hierarchical clustering methods. *Int. J. Remote Sens.* 29, 3171–3207. <https://doi.org/10.1080/01431160701442146>
- Gorte, B., Stein, A., 1998. Bayesian classification and class area estimation of satellite images using stratification. *IEEE Trans. Geosci. Remote Sens.* 36, 803–812. <https://doi.org/10.1109/36.673673>
- Government of Ethiopia, 2018. Ethiopia Government Portal - | Oromia Regional State [WWW Document]. URL <https://ictet.org/ethiopia/regional-states/oromia-regional-state/> (accessed 8.7.22).
- Grace, K., Husak, G.J., Harrison, L., Pedreros, D., Michaelsen, J., 2012. Using high resolution satellite imagery to estimate cropped area in Guatemala and Haiti. *Appl. Geogr.* 32, 433–440. <https://doi.org/10.1016/j.apgeog.2011.05.014>
- Han, W., Yang, Z., Di, L., Mueller, R., 2012. CropScape: A Web service based application for exploring and disseminating US conterminous geospatial cropland data products for decision support. *Comput. Electron. Agric.* 84, 111–123. <https://doi.org/10.1016/J.COMPAG.2012.03.005>
- Hird, J.N., McDermid, G.J., 2009. Noise reduction of NDVI time series: An empirical comparison of selected techniques. *Remote Sens. Environ.* 113, 248–258. <https://doi.org/10.1016/j.rse.2008.09.003>
- Huang, S., Tang, L., Hupy, J.P., Wang, Y., Shao, G., 2021. A commentary review on the use of normalized difference vegetation index (NDVI) in the era of popular remote sensing. *J. For. Res.* 32, 1–6. <https://doi.org/10.1007/S11676-020-01155-1/FIGURES/2>
- Husak, G., Grace, K., 2016. In search of a global model of cultivation: using remote sensing to examine the characteristics and constraints of agricultural production in the developing world. *Food Secur.* 8,

- 167–177. <https://doi.org/10.1007/S12571-015-0538-6/FIGURES/3>
- Husak, G.J., Marshall, M.T., Michaelsen, J., Pedreros, D., Funk, C., Galu, G., 2008. Crop area estimation using high and medium resolution satellite imagery in areas with complex topography. *J. Geophys. Res.* 113, D14112. <https://doi.org/10.1029/2007JD009175>
- Jackson, P.H., Ferguson, G.A., 1972. *Statistical Analysis in Psychology and Education*. *J. R. Stat. Soc. Ser. A* 135, 153. <https://doi.org/10.2307/2345050>
- Khan, M.R., de Bie, C.A.J.M., van Keulen, H., Smaling, E.M.A., Real, R., 2010. Disaggregating and mapping crop statistics using hypertemporal remote sensing. *Int. J. Appl. Earth Obs. Geoinf.* 12, 36–46. <https://doi.org/10.1016/j.jag.2009.09.010>
- Kvamme, K.L., Ernenwein, E.G., Menzer, J.G., 2019. Putting it all together: Geophysical data integration. *Innov. Near-Surface Geophys.* 287–339. <https://doi.org/10.1016/B978-0-12-812429-1.00009-X>
- Lesiv, M., Laso Bayas, J.C., See, L., Duerauer, M., Dahlia, D., Durando, N., Hazarika, R., Kumar Sahariah, P., Vakolyuk, M., Blyshchyk, V., Bilous, A., Perez-Hoyos, A., Gengler, S., Prestele, R., Bilous, S., Akhtar, I. ul H., Singha, K., Choudhury, S.B., Chetri, T., Malek, Ž., Bungnamei, K., Saikia, A., Sahariah, D., Narzary, W., Danylo, O., Sturn, T., Karner, M., McCallum, I., Schepaschenko, D., Moltchanova, E., Fraisl, D., Moorthy, I., Fritz, S., 2019. Estimating the global distribution of field size using crowdsourcing. *Glob. Chang. Biol.* 25, 174–186. <https://doi.org/10.1111/gcb.14492>
- Malladi, R.M.V., Nizami, A., Mahakali, M.S., Krishna, B.G., 2018. Cloud Masking Technique for High-Resolution Satellite Data: An Artificial Neural Network Classifier Using Spectral & Textural Context. *J. Indian Soc. Remote Sens.* 2018 474 47, 661–670. <https://doi.org/10.1007/S12524-018-0892-X>
- Marshall, M.T., Husak, G.J., Michaelsen, J., Funk, C., Pedreros, D., Adoum, A., 2011. Testing a high-resolution satellite interpretation technique for crop area monitoring in developing countries. *Int. J. Remote Sens.* 32, 7997–8012. <https://doi.org/10.1080/01431161.2010.532168>
- Maselli, F., Moriondo, M., Angeli, L., Fibbi, L., Bindi, M., 2011. Estimation of wheat production by the integration of MODIS and ground data. *Int. J. Remote Sens.* 32, 1105–1123. <https://doi.org/10.1080/01431160903510799>
- Meneses-Tovar, C.L., 2011. NDVI as indicator of degradation. *Unasylva* 62(238), 39–46. <https://doi.org/https://www.fao.org/3/i2560e/i2560e07.pdf>
- Mohamed, A.A., 2017. Food Security Situation in Ethiopia: A Review Study. *Int. J. Heal. Econ. Policy* 2, 86–96. <https://doi.org/10.11648/J.HEP.20170203.11>
- Mohammed, I., Marshall, M., de Bie, K., Estes, L., Nelson, A., 2020a. A blended census and multiscale remote sensing approach to probabilistic cropland mapping in complex landscapes. *ISPRS J. Photogramm. Remote Sens.* 161, 233–245. <https://doi.org/10.1016/j.isprsjprs.2020.01.024>
- Mohammed, I., Marshall, M., de Bie, K., Estes, L., Nelson, A., 2020b. A blended census and multiscale remote sensing approach to probabilistic cropland mapping in complex landscapes. *ISPRS J. Photogramm. Remote Sens.* 161, 233–245. <https://doi.org/10.1016/j.isprsjprs.2020.01.024>
- Natteshan, N.V.S., Suresh Kumar, N., 2020. Effective SAR image segmentation and classification of crop areas using MRG and CDNN techniques. *tnl homepage: https://www.tandfonline.com/loi/tejr20 Effective SAR image segmentation and classification of crop areas using MRG and CDNN techniques. Eur. J. Remote Sens.* 53, 126–140. <https://doi.org/10.1080/22797254.2020.1727777>
- Ochola, E.M., Fakharizadehshirazi, E., Adimo, A.O., Mukundi, J.B., Wesonga, J.M., Sodoudi, S., 2020. Inter-local climate zone differentiation of land surface temperatures for Management of Urban Heat in Nairobi City, Kenya. *Urban Clim.* 31. <https://doi.org/10.1016/j.uclim.2019.100540>
- Pan, Z., Huang, J., Zhou, Q., Wang, L., Cheng, Y., Zhang, H., Blackburn, G.A., Yan, J., Liu, J., 2015. Mapping crop phenology using NDVI time-series derived from HJ-1 A/B data. *Int. J. Appl. Earth Obs. Geoinf.* 34, 188–197. <https://doi.org/10.1016/j.jag.2014.08.011>
- Pandey, P., Kington, J., Kanwar, A., Curdoglo, M., 2021. Addendum to Planet Basemaps Product Specifications NICFI Basemaps. [https://doi.org/https://assets.planet.com/docs/NICFI\\_Basemap\\_Spec\\_Addendum.pdf](https://doi.org/https://assets.planet.com/docs/NICFI_Basemap_Spec_Addendum.pdf)
- PLANET, 2021. NICFI DATA Program User Guide Third party participants (Level 1 Users) 1–12.
- Planet Team, 2017. Planet Application Program Interface: In Space for Life on Earth. San Francisco, CA. [<https://api.planet.com>] [WWW Document]. URL [https://developers.google.com/earth-engine/datasets/catalog/projects\\_planet-nicfi\\_assets\\_basemaps\\_africa](https://developers.google.com/earth-engine/datasets/catalog/projects_planet-nicfi_assets_basemaps_africa) (accessed 5.19.22).
- Pradhan, S., 2001. Crop area estimation using GIS, remote sensing and area frame sampling. *Int. J. Appl. Earth Obs. Geoinf.* 3, 86–92. [https://doi.org/10.1016/S0303-2434\(01\)85025-X](https://doi.org/10.1016/S0303-2434(01)85025-X)
- Recio, J.A., Hermosilla, T., Ruiz, L., Fernández-Sarría, A., 2010. Addition of geographic ancillary data for updating geo-spatial databases. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. - ISPRS Arch.*

- Ruxton, G.D., Beauchamp, G., 2008. Time for some a priori thinking about post hoc testing. *Behav. Ecol.* 19, 690–693. <https://doi.org/10.1093/BEHECO/ARN020>
- Savitzky, A., Golay, M.J., 1964. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Am. Pharm. Assoc., Sci. Ed* 36, 1627–1639. <https://doi.org/1964AnaCh..36.1627S>
- Shao, Y., Fan, X., Liu, H., Xiao, J., Ross, S., Brisco, B., Brown, R., Staples, G., 2001. Rice monitoring and production estimation using multitemporal RADARSAT. *Remote Sens. Environ.* 76, 310–325. [https://doi.org/10.1016/S0034-4257\(00\)00212-1](https://doi.org/10.1016/S0034-4257(00)00212-1)
- Shih, J.H., Fay, M.P., 2017. Pearson's chi-square test and rank correlation inferences for clustered data. *Biometrics* 73, 822–834. <https://doi.org/10.1111/BIOM.12653>
- Sud, U.C., Ahmad, T., Gupta, V, K., Chandra, H., Sahoo, P.M., Aditya, K., Man, S., Biswas, A., 2017. Methodology for Estimation of Crop Area and Crop Yield under Mixed and Continuous Cropping. FAO, Rome Publication. <https://doi.org/https://www.fao.org/3/ca6514en/ca6514en.pdf>
- Taylor, J.C., Waine, T.W., Juniper, G.R., Simms, D.M., Brewer, T.R., 2010. Survey and monitoring of opium poppy and wheat in Afghanistan: 2003-2009. *Remote Sens. Lett.* 179–185.
- The World Bank, 2008. World development report, 2008: agriculture for development. *Choice Rev. Online* 45, 45-4765-45–4765. <https://doi.org/10.5860/choice.45-4765>
- UDA Consulting, 2018. An Overview of Agricultural Statistical Systems in the Selected Countries. Ankara, Turkiye 0–30.
- USAID, 2020. Agriculture and Food Security | Ethiopia | U.S. Agency for International Development [WWW Document]. URL <https://www.usaid.gov/ethiopia/agriculture-and-food-security> (accessed 11.20.21).
- Veloso, A., Mermoz, S., Bouvet, A., Le Toan, T., Planells, M., Dejoux, J.F., Ceschia, E., 2017. Understanding the temporal behavior of crops using Sentinel-1 and Sentinel-2-like data for agricultural applications. *Remote Sens. Environ.* 199, 415–426. <https://doi.org/10.1016/J.RSE.2017.07.015>
- Vintrou, E., Desbrosse, A., Bégue, A., Traoré, S., Baron, C., Lo Seen, D., 2012. Crop area mapping in West Africa using landscape stratification of MODIS time series and comparison with existing global land products. *Int. J. Appl. Earth Obs. Geoinf.* 14, 83–93. <https://doi.org/10.1016/j.jag.2011.06.010>
- Wang, D., LI, Z.L., ZHOU, Q.B., YANG, P., CHEN, Z.X., 2018. An optimized two-stage spatial sampling scheme for winter wheat acreage estimation using remotely sensed imagery. *Int. J. Remote Sens.* 40, 2014–2033. <https://doi.org/10.1080/01431161.2018.1516321>
- Werner, J.P.S., Oliveira, S.R.D.M., Esquerdo, J.C.D.M., 2019. Mapping cotton fields using data mining and MODIS time-series. *Int. J. Remote Sens.* 41, 2457–2476. <https://doi.org/10.1080/01431161.2019.1693072>
- WFP, CSA, 2019. Comprehensive Food Security and Vulnerability Analysis (CFSVA). <https://doi.org/https://docs.wfp.org/api/documents/WFP-0000106754/download/>
- World Bank, 2011. GLOBAL STRATEGY TO IMPROVE AGRICULTURAL AND RURAL STATISTICS United Nations. Washington DC.
- World Bank, 2010. Global Strategy To Improve Agricultural and Rural Statistics, The World Bank, Washington DC. <https://doi.org/https://openknowledge.worldbank.org/handle/10986/12402?locale-attribute=fr>

# ANNEXES

## Annex-1. GEE Code used to compute and export the time series NDVI of the 400 tiles.

```
1 // This code used to compute and export the time series NDVI of NICFI funded PlanetScope imagery
2 // sampleplots are shape files of the sample area(tiles)
3 // geometry is to show the study area
4 //=====
5 // Load Geometry
6 //=====
7 var nicfi = ee.ImageCollection('projects/planet-nicfi/assets/basemaps/africa');
8 var samplePlots = ee.FeatureCollection("projects/itcshafi/assets/Selected_Sample_polygon");
9 var geometry = ee.FeatureCollection("projects/itcshafi/assets/Selected_Sample_polygon");
10
11 //=====
12 // Create Time Series (TS) of monthly composites
13 //=====
14
15 var TS = ee.ImageCollection('projects/planet-nicfi/assets/basemaps/africa')
16     .filterDate('2021-01-01','2021-12-31')
17     .filterMetadata('cadence', 'equals','monthly')
18     .filterBounds(geometry)
19     .map(function(image){return image.clip(geometry)});
20
21 var year = TS.median().clip(geometry);
22
23 //=====
24 //===== Individual Month cloud Masking=====
25 // Masking clouds of April
26 //=====
27 var month = ee.ImageCollection('projects/planet-nicfi/assets/basemaps/africa')
28     .filterDate('2021-04-01','2021-04-30')
29     .filterMetadata('cadence', 'equals','monthly');
30
31 var month = month.median().clip(geometry);
32 var cloudMaskApr = month.select('G').subtract(year.select('G')).lt(700);
33 var Apr_masked = month.updateMask(cloudMaskApr);
34
35 //=====
36 // Masking clouds of May
37 //=====
38 var month = ee.ImageCollection('projects/planet-nicfi/assets/basemaps/africa')
39     .filterDate('2021-05-01','2021-05-30')
40     .filterMetadata('cadence', 'equals','monthly');
41
42 var month = month.median().clip(geometry);
43 var cloudMaskMay = month.select('G').subtract(year.select('G')).lt(700);
44 var May_masked = month.updateMask(cloudMaskMay);
45
46 //=====
47 // Masking clouds Jun
48 //=====
49 var month = ee.ImageCollection('projects/planet-nicfi/assets/basemaps/africa')
50     .filterDate('2021-06-01','2021-06-30')
51     .filterMetadata('cadence', 'equals','monthly');
52
53 var month = month.median().clip(geometry);
54 var cloudMaskJun = month.select('G').subtract(year.select('G')).lt(700);
55 var Jun_masked = month.updateMask(cloudMaskJun);
56
57 //=====
58 // Masking clouds of Jul
59 //=====
60 var month = ee.ImageCollection('projects/planet-nicfi/assets/basemaps/africa')
61     .filterDate('2021-07-01','2021-07-30')
62     .filterMetadata('cadence', 'equals','monthly');
63
64 var month = month.median().clip(geometry);
65 var cloudMaskJul = month.select('G').subtract(year.select('G')).lt(400);
66 var Jul_masked = month.updateMask(cloudMaskJul);
67
68 //=====
69 // Masking clouds of Aug
70 //=====
```

```

68 //=====
69 // Masking clouds of Aug
70 //=====
71 var month = ee.ImageCollection('projects/planet-nicfi/assets/basemaps/africa')
72     .filterDate('2021-08-01','2021-08-30')
73     .filterMetadata('cadence', 'equals','monthly');
74
75 var month = month.median().clip(geometry);
76 var cloudMaskAug = month.select('G').subtract(year.select('G')).lt(500);
77 var Aug_masked = month.updateMask(cloudMaskAug);
78
79 //=====
80 // Masking clouds of Sep
81 //=====
82 var month = ee.ImageCollection('projects/planet-nicfi/assets/basemaps/africa')
83     .filterDate('2021-09-01','2021-09-30')
84     .filterMetadata('cadence', 'equals','monthly');
85
86 var month = month.median().clip(geometry);
87 var cloudMaskSep = month.select('G').subtract(year.select('G')).lt(500);
88 var Sep_masked = month.updateMask(cloudMaskSep);
89
90 //=====
91 // Masking clouds of Oct
92 //=====
93 var month = ee.ImageCollection('projects/planet-nicfi/assets/basemaps/africa')
94     .filterDate('2021-10-01','2021-10-30')
95     .filterMetadata('cadence', 'equals','monthly');
96
97 var month = month.median().clip(geometry);
98 var cloudMaskOct = month.select('G').subtract(year.select('G')).lt(700);
99 var Oct_masked = month.updateMask(cloudMaskOct);
100
101 //=====
102 // Create TS of monthly composites masked for some month
103 //=====
104
105 var Jan = ee.ImageCollection('projects/planet-nicfi/assets/basemaps/africa')
106     .filterDate('2021-01-01','2021-01-31')
107     .filterMetadata('cadence', 'equals','monthly')
108     .filterBounds(geometry)
109     .map(function(image){return image.clip(geometry)});
110 var Feb = ee.ImageCollection('projects/planet-nicfi/assets/basemaps/africa')
111     .filterDate('2021-02-01','2021-02-28')
112     .filterMetadata('cadence', 'equals','monthly')
113     .filterBounds(geometry)
114     .map(function(image){return image.clip(geometry)});
115 var Mar = ee.ImageCollection('projects/planet-nicfi/assets/basemaps/africa')
116     .filterDate('2021-03-01','2021-03-31')
117     .filterMetadata('cadence', 'equals','monthly')
118     .filterBounds(geometry)
119     .map(function(image){return image.clip(geometry)});
120
121 var Nov = ee.ImageCollection('projects/planet-nicfi/assets/basemaps/africa')
122     .filterDate('2021-11-01','2021-11-30')
123     .filterMetadata('cadence', 'equals','monthly')
124     .filterBounds(geometry)
125     .map(function(image){return image.clip(geometry)});
126 var Dec = ee.ImageCollection('projects/planet-nicfi/assets/basemaps/africa')
127     .filterDate('2021-12-01','2021-12-30')
128     .filterMetadata('cadence', 'equals','monthly')
129     .filterBounds(geometry)
130     .map(function(image){return image.clip(geometry)});
131
132
133 var Jan = Jan.median().clip(geometry);
134 var Feb = Feb.median().clip(geometry);
135 var Mar = Mar.median().clip(geometry);
136 var Apr = Apr_masked.clip(geometry);
137 var May = May_masked.clip(geometry);
138 var Jun = Jun_masked.clip(geometry);

```

```

139 var Jul = Jul_masked.clip(geometry);
140 var Aug = Aug_masked.clip(geometry);
141 var Sep = Sep_masked.clip(geometry);
142 var Oct = Oct_masked.clip(geometry);
143 var Nov = Nov.median().clip(geometry);
144 var Dec = Dec.median().clip(geometry);
145
146 var TS_image = ee.Image.cat([ Jan, Feb, Mar, Apr, May, Jun, Jul ,Aug, Sep, Oct, Nov, Dec]);
147 // =====
148 // creating functions to export all sample area(400X12)
149 // =====
150 for (var i=1; i<2; i++) {
151
152 var geometry = samplePLots.filter(ee.Filter.eq("ID", i));
153 // =====
154 // NDVI processing and data Streching
155 // =====
156 var Jan_NDVI = Jan.normalizedDifference(['N', 'R']).unitScale(-1, 1).multiply(255).toByte();
157 var Feb_NDVI = Feb.normalizedDifference(['N', 'R']).unitScale(-1, 1).multiply(255).toByte();
158 var Mar_NDVI = Mar.normalizedDifference(['N', 'R']).unitScale(-1, 1).multiply(255).toByte();
159 var Apr_NDVI = Apr.normalizedDifference(['N', 'R']).unitScale(-1, 1).multiply(255).toByte();
160 var May_NDVI = May.normalizedDifference(['N', 'R']).unitScale(-1, 1).multiply(255).toByte();
161 var Jun_NDVI = Jun.normalizedDifference(['N', 'R']).unitScale(-1, 1).multiply(255).toByte();
162 var Jul_NDVI = Jul.normalizedDifference(['N', 'R']).unitScale(-1, 1).multiply(255).toByte();
163 var Aug_NDVI = Aug.normalizedDifference(['N', 'R']).unitScale(-1, 1).multiply(255).toByte();
164 var Sep_NDVI = Sep.normalizedDifference(['N', 'R']).unitScale(-1, 1).multiply(255).toByte();
165 var Oct_NDVI = Oct.normalizedDifference(['N', 'R']).unitScale(-1, 1).multiply(255).toByte();
166 var Nov_NDVI = Nov.normalizedDifference(['N', 'R']).unitScale(-1, 1).multiply(255).toByte();
167 var Dec_NDVI =Dec.normalizedDifference(['N', 'R']).unitScale(-1, 1).multiply(255).toByte();
168 // =====
169 // Setting 0 for the the missing value
170 // =====
171 var Jan_NDVI = Jan_NDVI.unmask(0).clip(geometry);
172 var Feb_NDVI = Feb_NDVI.unmask(0).clip(geometry);
173 var Mar_NDVI = Mar_NDVI.unmask(0).clip(geometry);
174 var Apr_NDVI = Apr_NDVI.unmask(0).clip(geometry);
175 var May_NDVI = May_NDVI.unmask(0).clip(geometry);
176 var Jun_NDVI = Jun_NDVI.unmask(0).clip(geometry);
177 var Jul_NDVI = Jul_NDVI.unmask(0).clip(geometry);
178 var Aug_NDVI = Aug_NDVI.unmask(0).clip(geometry);
179 var Sep_NDVI = Sep_NDVI.unmask(0).clip(geometry);
180 var Oct_NDVI = Oct_NDVI.unmask(0).clip(geometry);
181 var Nov_NDVI = Nov_NDVI.unmask(0).clip(geometry);
182 var Dec_NDVI = Dec_NDVI.unmask(0).clip(geometry);
183 // =====
184 // NDVI Time series Visualization
185 // =====
186 var NDVI_TS = ee.Image.cat([ Jan_NDVI, Feb_NDVI, Mar_NDVI, Apr_NDVI, May_NDVI,
187 Jun_NDVI, Jul_NDVI ,Aug_NDVI, Sep_NDVI, Oct_NDVI, Nov_NDVI, Dec_NDVI]);
188 var vis = {"bands":["R","G","B"],"min":64,"max":5454,"gamma":1.8};
189 Map.addLayer(NDVI_TS.randomVisualizer(), {}, 'NDVI');
190 Map.addLayer(TS_image.randomVisualizer(), {}, 'Image');
191 // =====
192 // Export time serie NDVI Data
193 // =====
194 Export.image.toDrive({
195   image: NDVI_TS,
196   description: 'NDVI_TS'+i,
197   folder: 'Tiles_TS',
198   region: geometry,
199   scale: 4.7
200 });
201
202 }
203

```

## Annex-2

### Description of The NDVI Profiles through Google Earth imagery inspection

Description				
	Google Earth imagery inspection			NDVI profiles
Custers	Mostly	sometimes	Almost none	Profile analysis
Class_1	Bare land	water bodies	crop field	Bare land
Class_2	Cropped field	Grass		Cropped field
Class_3	Cropped field		Grass	Cropped field
Class_4	Bare lands		Bare lands	None-cropped
Class_5	Cropped field			cropped field
Class_6	Cropped field			cropped field
Class_7	Grass land	cropped field		Grass land
Class_8	Cropped field			cropped field
Class_9	Cropped field	Grass land		cropped field
Class_10	Cropped field	Grass land		cropped field
Class_11	Cropped field			cropped field
Class_12	Shrubs land(trees)	grass		shrub lands
Class_13	Grassland	cropped field		Grass land
Class_14	Grassland			Grass Land
Class_15	Shrubs land(trees)	trees	Cropped field	shrubs land
Class_16	Shrubs land (Tress)	cropped field	Trees	Shrubs land
Class_17	Grassland		cropped field	
Class_18	Shrubs land	Grass	Tree(single)	shrubs land
Class_19	Grassland	shrubs	cropped field	Grass Land
Class_20	Forest (Trees dense)		Grass (Swampy area)	Trees(dense)
Class_21	Forest (Trees Dense)			Trees(dense)
Class_22	Forest (Trees Dense)			Trees(dense)
Class_23	Forest (Trees Dense)			Trees(dense)
Class_24	Forest (Trees Dense)			Trees(dense)
Class_25	Forest (Trees Dense)			Trees(dense)

### Annex-3

The average area fraction of the grouped NDVI clusters based on spectral similarity group according to the Strata maps

Mean area fraction of the grouped NDVI clusters according to the Strata maps											
CPS zone						Woreda					
CPS zone	Base land	Crop fields	Grasslands	Trees	Forest	W_Name	Base land	Crop fields	Grass lands	Trees	Forest
CPS_19	0.73	9.88	28.59	31.54	29.8	Abaya	0	0	22.75	7.87	69.75
CPS_20	3.9	26.45	26.26	31.64	12.3	Adea	18.25	68.13	9	4.88	0
CPS_29	36.13	44.53	11.28	8.22	0.38	Adea Berga	1.88	17.25	38.88	22.75	20.25
CPS_30	32	40.37	23.38	4.56	0	Ambo	3.25	26.31	19.94	35.19	15.75
CPS_31	19.06	69.06	8.19	3.94	0	Ameya	5.5	51.13	23.88	16.38	3.75
CPS_33	6.98	48.83	27.14	15.97	1.75	Amigna	0.38	47.75	36.13	16.13	0.38
CPS_34	4.66	20.72	27.75	43.78	3.47	Anfilo	0	0	0	0	100.63
CPS_35	2.25	26.62	37.63	31.63	2.38	Arsi Negele	0.25	41.37	32	14.88	11.63
CPS_43	0.46	15.75	14.79	58.63	10.83	Ayia	0	15.38	33.75	17.63	33.88
CPS_48	0	0	0	0	100.63	Bale Gasera	2.25	26.62	37.63	31.63	2.38
CPS_54	0.04	0	15.96	7.76	76.6	Bosa	19.88	70	7.38	3	0
CPS_57	0.03	3.77	19.95	5.56	71.26	Boset	1.5	7.37	28.25	62.63	0.13
						Chelia	13.25	42.25	24.88	16	3.62
						Dawo	35.13	51.63	7.63	5.5	0.25
						Dendi	39.25	37.13	16.38	7.87	0.13
						Diga	0	0.13	31.38	22.5	46.63
						Digehuna Ti	4	70.25	16.75	8.88	0.75
						Dodola	0.13	0	6	1.75	93.38
						Dugdadewa	0	0	11.88	14.38	74.25
						Gemechis	0.13	0.63	27.25	32.63	40.13
						GINNIR	0.88	46.88	31.75	19.13	1.63
						Girawa	0.75	7.25	33.25	47.13	12
						Gole Oda	0	1.13	15.88	65.5	17.63
						Guba Qonic	4.63	19.5	23.38	38.62	13.75
						Horo	0.5	38.13	29.5	21.63	10.63
						Ilu	47.13	36.13	7.25	9.25	0.5
						Leqa Dulech	0	1.25	37.13	26.25	35.63
						Liben	0.25	0.13	2.5	87.63	10
						Limuna Bill	7.75	67.38	16.63	8.38	0.5
						Meda Welab	0.25	0.25	10.13	69.13	20.88
						Meiso	34.63	22.5	38.25	5.13	0
						Merti	6.75	29.88	27.88	32	4.13
						Meta	2.13	17.5	31.5	44	6
						Metarobi	13.5	24.06	24.13	34.75	4.38
						Mulo	9.38	41.25	27.63	16.5	5.5
						Odo Shakis	0.13	0	22.38	2.63	75.88
						Omonada	0.13	7.75	24	43.62	25
						Robe	0	54.25	31.88	13	2.25
						Seyo Nole	0	2.88	31.38	3.75	61.5
						Sude	0	7	31.88	29.88	32.12
						Tiyo	4.13	47.25	24.5	22.75	1.88
						Uraga	0	0.06	9.5	1.69	89.06
						Weliso	2.62	37.62	28.13	23.5	10
						Welmera	23	53.25	13.88	10.25	0.63
						Wuchale	14.88	31.63	36.88	16	1.25
						Zeway Dugi	29.38	58.25	8.5	4	0

## Annex-4

The significantly different pairs of strata of the crop field cluster of the three stratification maps

### A) CPS zone Strata map

Multiple Comparisons

Dependent Variable: Crop fields clusters

Hochberg

(I) CPS Zone	(J) CPS Zone	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval		
					Lower Bound	Upper Bound	
19	20	-16.575*	2.966	0.001	-26.62	-6.53	
	29	-34.656*	3.772	0.001	-47.43	-21.88	
	30	-30.500*	4.826	0.001	-46.84	-14.16	
	31	-59.187*	4.826	0.001	-75.53	-42.85	
	33	-38.953*	3.115	0.001	-49.5	-28.41	
	34	-10.844	3.772	0.243	-23.62	1.93	
	35	-16.75	6.434	0.463	-38.54	5.04	
	43	-5.875	4.153	1	-19.94	8.19	
	48	9.875	6.434	1	-11.91	31.66	
	54	9.875	4.095	0.652	-3.99	23.74	
	57	6.106	3.55	0.996	-5.92	18.13	
	20	19	16.575*	2.966	0.001	6.53	26.62
		29	-18.081*	3.561	0.001	-30.14	-6.02
		30	-13.925	4.662	0.178	-29.71	1.86
31		-42.612*	4.662	0.001	-58.4	-26.83	
33		-22.378*	2.855	0.001	-32.05	-12.71	
34		5.731	3.561	0.999	-6.33	17.79	
35		-0.175	6.312	1	-21.55	21.2	
43		10.7	3.962	0.374	-2.72	24.12	
48		26.450*	6.312	0.002	5.08	47.82	
54		26.450*	3.9	0.001	13.24	39.66	
57		22.681*	3.325	0.001	11.42	33.94	
29		19	34.656*	3.772	0.001	21.88	47.43
		20	18.081*	3.561	0.001	6.02	30.14
		30	4.156	5.212	1	-13.49	21.81
	31	-24.531*	5.212	0.001	-42.18	-6.88	
	33	-4.297	3.686	1	-16.78	8.18	
	34	23.813*	4.256	0.001	9.4	38.22	
	35	17.906	6.729	0.409	-4.88	40.69	
	43	28.781*	4.597	0.001	13.22	44.35	
	48	44.531*	6.729	0.001	21.75	67.32	
	54	44.531*	4.544	0.001	29.14	59.92	
	57	40.762*	4.06	0.001	27.01	54.51	
	30	19	30.500*	4.826	0.001	14.16	46.84
		20	13.925	4.662	0.178	-1.86	29.71
		29	-4.156	5.212	1	-21.81	13.49
31		-28.687*	6.018	0.001	-49.07	-8.31	
33		-8.453	4.758	0.993	-24.57	7.66	
34		19.656*	5.212	0.012	2.01	37.31	
35		13.75	7.371	0.983	-11.21	38.71	
43		24.625*	5.494	0.001	6.02	43.23	
48		40.375*	7.371	0.001	15.41	65.34	
54		40.375*	5.45	0.001	21.92	58.83	
57		36.606*	5.054	0.001	19.49	53.72	
31		19	59.188*	4.826	0.001	42.85	75.53
		20	42.613*	4.662	0.001	26.83	58.4
		29	24.531*	5.212	0.001	6.88	42.18
	30	28.688*	6.018	0.001	8.31	49.07	
	33	20.234*	4.758	0.002	4.12	36.35	
	34	48.344*	5.212	0.001	30.69	65.99	
	35	42.438*	7.371	0.001	17.48	67.4	
	43	53.313*	5.494	0.001	34.71	71.92	
	48	69.063*	7.371	0.001	44.1	94.02	
	54	69.063*	5.45	0.001	50.61	87.52	
	57	65.293*	5.054	0.001	48.18	82.41	
	33	19	38.953*	3.115	0.001	28.41	49.5
		20	22.378*	2.855	0.001	12.71	32.05
		29	4.297	3.686	1	-8.18	16.78
30		8.453	4.758	0.993	-7.66	24.57	
31		-20.234*	4.758	0.002	-36.35	-4.12	
34		28.109*	3.686	0.001	15.63	40.59	
35		22.203*	6.384	0.036	0.59	43.82	
43		33.078*	4.075	0.001	19.28	46.88	
48		48.828*	6.384	0.001	27.21	70.44	
54		48.828*	4.015	0.001	35.23	62.42	

	57	45.059*	3.458	0.001	33.35	56.77
<b>34</b>	19	10.844	3.772	0.243	-1.93	23.62
	20	-5.731	3.561	0.999	-17.79	6.33
	29	-23.812*	4.256	0.001	-38.22	-9.4
	30	-19.656*	5.212	0.012	-37.31	-2.01
	31	-48.344*	5.212	0.001	-65.99	-30.69
	33	-28.109*	3.686	0.001	-40.59	-15.63
	35	-5.906	6.729	1	-28.69	16.88
	43	4.969	4.597	1	-10.6	20.53
	48	20.719	6.729	0.135	-2.07	43.5
	54	20.719*	4.544	0.001	5.33	36.11
	57	16.950*	4.06	0.002	3.2	30.7
<b>35</b>	19	16.75	6.434	0.463	-5.04	38.54
	20	0.175	6.312	1	-21.2	21.55
	29	-17.906	6.729	0.409	-40.69	4.88
	30	-13.75	7.371	0.983	-38.71	11.21
	31	-42.437*	7.371	0.001	-67.4	-17.48
	33	-22.203*	6.384	0.036	-43.82	-0.59
	34	5.906	6.729	1	-16.88	28.69
	43	10.875	6.95	1	-12.66	34.41
	48	26.625	8.511	0.116	-2.2	55.45
	54	26.625*	6.915	0.009	3.21	50.04
	57	22.856*	6.607	0.039	0.48	45.23
<b>43</b>	19	5.875	4.153	1	-8.19	19.94
	20	-10.7	3.962	0.574	-24.12	2.72
	29	-28.781*	4.597	0.001	-44.35	-13.22
	30	-24.625*	5.494	0.001	-43.23	-6.02
	31	-53.312*	5.494	0.001	-71.92	-34.71
	33	-33.078*	4.075	0.001	-46.88	-19.28
	34	-4.969	4.597	1	-20.53	10.6
	35	-10.875	6.95	1	-34.41	12.66
	48	15.75	6.95	0.787	-7.78	39.28
	54	15.75	4.865	0.082	-0.72	32.22
	57	11.981	4.416	0.364	-2.97	26.94
<b>48</b>	19	-9.875	6.434	1	-31.66	11.91
	20	-26.450*	6.312	0.002	-47.82	-5.08
	29	-44.531*	6.729	0.001	-67.32	-21.75
	30	-40.375*	7.371	0.001	-65.34	-15.41
	31	-69.062*	7.371	0.001	-94.02	-44.1
	33	-48.828*	6.384	0.001	-70.44	-27.21
	34	-20.719	6.729	0.135	-43.5	2.07
	35	-26.625	8.511	0.116	-55.45	2.2
	43	-15.75	6.95	0.787	-39.28	7.78
	54	0	6.915	1	-23.42	23.42
	57	-3.769	6.607	1	-26.14	18.6
<b>54</b>	19	-9.875	4.095	0.652	-23.74	3.99
	20	-26.450*	3.9	0.001	-39.66	-13.24
	29	-44.531*	4.544	0.001	-59.92	-29.14
	30	-40.375*	5.45	0.001	-58.83	-21.92
	31	-69.062*	5.45	0.001	-87.52	-50.61
	33	-48.828*	4.015	0.001	-62.42	-35.23
	34	-20.719*	4.544	0.001	-36.11	-5.33
	35	-26.625*	6.915	0.009	-50.04	-3.21
	43	-15.75	4.865	0.082	-32.22	0.72
	48	0	6.915	1	-23.42	23.42
	57	-3.769	4.361	1	-18.54	11
<b>57</b>	19	-6.106	3.55	0.996	-18.13	5.92
	20	-22.681*	3.325	0.001	-33.94	-11.42
	29	-40.762*	4.06	0.001	-54.51	-27.01
	30	-36.606*	5.054	0.001	-53.72	-19.49
	31	-65.293*	5.054	0.001	-82.41	-48.18
	33	-45.059*	3.458	0.001	-56.77	-33.35
	34	-16.950*	4.06	0.002	-30.7	-3.2
	35	-22.856*	6.607	0.039	-45.23	-0.48
	43	-11.981	4.416	0.364	-26.94	2.97
	48	3.769	6.607	1	-18.6	26.14
	54	3.769	4.361	1	-11	18.54

\* The mean difference is significant at the 0.05 level.

## B) LH Zone Map

The sig. value copied here only for some pairs of strata because the table is large.

Multiple Comparisons

Dependent Variable: Crop field clusters						
Hochberg						
(I) LHZ_No	(J) LHZ_No	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
2	4	43.163*	6.27	0.001	19.8	66.53
	13	6.225	4.631	1	-11.03	23.48
	21	-1.619	3.537	1	-14.8	11.56
	24	-8.796	3.935	0.995	-23.46	5.87
	31	43.131*	3.537	0	29.95	56.31
	35	42.913*	6.27	0.001	19.55	66.28
	37	39.225*	4.631	0.001	21.97	56.48
	41	4.788	6.27	1	-18.58	28.15
	48	27.788*	6.27	0.003	4.42	51.15
	50	41.913*	6.27	0.001	18.55	65.28
	80	43.038*	6.27	0.001	19.67	66.4
	100	22.371*	3.935	0.001	7.71	37.04
	113	20.663	6.27	0.216	-2.7	44.03
	117	35.413*	6.27	0.001	12.05	58.78
	128	43.100*	4.631	0	25.84	60.36
	134	36.163*	6.27	0.001	12.8	59.53
	136	-2.046	3.935	1	-16.71	12.62
138	13.913	4.631	0.468	-3.34	31.17	
148	28.756*	3.537	0.001	15.58	41.94	
169	5.038	6.27	1	-18.33	28.4	
174	40.288*	6.27	0.001	16.92	63.65	
4	2	-43.162*	6.27	0.001	-66.53	-19.8
	13	-36.937*	7.322	0.001	-64.22	-9.65
	21	-44.781*	6.684	0.001	-69.69	-19.87
	24	-51.958*	6.903	0.001	-77.68	-26.23
	31	-0.031	6.684	1	-24.94	24.88
	35	-0.25	8.455	1	-31.76	31.26
	37	-3.937	7.322	1	-31.22	23.35
	41	-38.375*	8.455	0.002	-69.88	-6.87
	48	-15.375	8.455	1	-46.88	16.13
	50	-1.25	8.455	1	-32.76	30.26
	80	-0.125	8.455	1	-31.63	31.38
	100	-20.792	6.903	0.46	-46.52	4.93
	113	-22.5	8.455	0.827	-54.01	9.01
	117	-7.75	8.455	1	-39.26	23.76
	128	-0.062	7.322	1	-27.35	27.22
	134	-7	8.455	1	-38.51	24.51
	136	-45.208*	6.903	0.001	-70.93	-19.48
138	-29.250*	7.322	0.018	-56.54	-1.96	
148	-14.406	6.684	0.999	-39.31	10.5	
169	-38.125*	8.455	0.002	-69.63	-6.62	
174	-2.875	8.455	1	-34.38	28.63	
13	2	-6.225	4.631	1	-23.48	11.03

	4	36.938*	7.322	0.001	9.65	64.22
	21	-7.844	5.177	1	-27.14	11.45
	24	-15.021	5.457	0.741	-35.36	5.32
	31	36.906*	5.177	0.001	17.61	56.2
	35	36.688*	7.322	0.001	9.4	63.97
	37	33.000*	5.978	0.001	10.72	55.28
	41	-1.437	7.322	1	-28.72	25.85
	48	21.563	7.322	0.532	-5.72	48.85
	50	35.688*	7.322	0.001	8.4	62.97
	80	36.813*	7.322	0.001	9.53	64.1
	100	16.146	5.457	0.517	-4.19	36.48
	113	14.438	7.322	1	-12.85	41.72
	117	29.188*	7.322	0.018	1.9	56.47
	128	36.875*	5.978	0.001	14.6	59.15
	134	29.938*	7.322	0.012	2.65	57.22
	136	-8.271	5.457	1	-28.61	12.07
	138	7.688	5.978	1	-14.59	29.97
	148	22.531*	5.177	0.004	3.24	41.82
	169	-1.187	7.322	1	-28.47	26.1
	174	34.063*	7.322	0.001	6.78	61.35
<b>21</b>	2	1.619	3.537	1	-11.56	14.8
	4	44.781*	6.684	0.001	19.87	69.69
	13	7.844	5.177	1	-11.45	27.14
	24	-7.177	4.566	1	-24.19	9.84
	31	44.750*	4.227	0	29	60.5
	35	44.531*	6.684	0.001	19.62	69.44
	37	40.844*	5.177	0.001	21.55	60.14
	41	6.406	6.684	1	-18.5	31.31
	48	29.406*	6.684	0.003	4.5	54.31
	50	43.531*	6.684	0.001	18.62	68.44
	80	44.656*	6.684	0.001	19.75	69.56
	100	23.990*	4.566	0.001	6.97	41.01
	113	22.281	6.684	0.192	-2.63	47.19
	117	37.031*	6.684	0.001	12.12	61.94
	128	44.719*	5.177	0	25.43	64.01
	134	37.781*	6.684	0.001	12.87	62.69
	136	-0.427	4.566	1	-17.44	16.59
	138	15.531	5.177	0.472	-3.76	34.82
	148	30.375*	4.227	0.001	14.62	46.13
	169	6.656	6.684	1	-18.25	31.56

\* The mean difference is significant at the 0.05 level.

### C) Woreda Strata map

Woreda names written under W\_code column with their woreda code. Because the table is large only some pairs of strata Sig. value copied here.

#### Multiple Comparisons

Dependent Variable: Crop field clusters						
(I) W_Code	(J) W_Code	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
<b>40118</b>	40120	-12.5	6	0.995	-36.44	11.44
	40210	2.75	6	1	-21.19	26.69
	40212	1.625	6	1	-22.31	25.56
	40411	-4.875	6	1	-28.81	19.06
	40503	-23.437*	5.196	0.007	-44.17	-2.71
	40505	-39.375*	6	0.001	-63.31	-15.44
	40510	-34.250*	6	0.001	-58.19	-10.31
	40512	-50.375*	6	0.001	-74.31	-26.44
	40513	-14.375	6	0.955	-38.31	9.56
	40514	-21.187*	5.196	0.037	-41.92	-0.46
	40608	-28.750*	6	0.002	-52.69	-4.81
	40616	-38.375*	6	0.001	-62.31	-14.44
	40702	-4.5	6	1	-28.44	19.44
	40706	-65.250*	6	0.001	-89.19	-41.31
	40710	-67.125*	6	0.001	-91.06	-43.19
	40711	2.75	6	1	-21.19	26.69
	40801	-27.000*	6	0.007	-50.94	-3.06
	40806	-55.375*	6	0.001	-79.31	-31.44
	40808	-4.125	6	1	-28.06	19.81
	40810	-44.875*	6	0.001	-68.81	-20.94
	40812	-51.375*	6	0.001	-75.31	-27.44
	40815	-67.375*	6	0.001	-91.31	-43.44
	40816	-44.375*	6	0.001	-68.31	-20.44
	40818	-64.500*	6	0.001	-88.44	-40.56
	40823	-23.75	6	0.056	-47.69	0.19
	40901	-19.625	6	0.366	-43.56	4.31
	40907	-16.625	5.196	0.422	-37.36	4.11
	40912	2.25	6	1	-21.69	26.19
	41009	-14.625	6	0.944	-38.56	9.31
	41016	-4.375	6	1	-28.31	19.56
41017	1.75	6	1	-22.19	25.69	
41108	-44.000*	6	0.001	-67.94	-20.06	
41113	2.625	6	1	-21.31	26.56	
41213	2.875	6	1	-21.06	26.81	
41215	2.875	6	1	-21.06	26.81	
41301	-48.250*	6	0.001	-72.19	-24.31	
41303	-34.750*	6	0.001	-58.69	-10.81	
41304	-48.750*	6	0.001	-72.69	-24.81	
41305	-33.250*	6	0.001	-57.19	-9.31	
41401	2.813	5.196	1	-17.92	23.54	
41405	2.875	6	1	-21.06	26.81	
41703	-38.500*	6	0.001	-62.44	-14.56	
41707	2.875	6	1	-21.06	26.81	
41807	2.875	6	1	-21.06	26.81	
41901	-35.250*	6	0.001	-59.19	-11.31	
<b>40120</b>	40118	12.5	6	0.995	-11.44	36.44
	40210	15.25	6	0.905	-8.69	39.19
	40212	14.125	6	0.965	-9.81	38.06
	40411	7.625	6	1	-16.31	31.56
	40503	-10.937	5.196	0.994	-31.67	9.79
	40505	-26.875*	6	0.008	-50.81	-2.94
	40510	-21.75	6	0.155	-45.69	2.19
	40512	-37.875*	6	0.001	-61.81	-13.94
	40513	-1.875	6	1	-25.81	22.06
	40514	-8.687	5.196	1	-29.42	12.04
	40608	-16.25	6	0.815	-40.19	7.69
	40616	-25.875*	6	0.015	-49.81	-1.94
	40702	8	6	1	-15.94	31.94
	40706	-52.750*	6	0.001	-76.69	-28.81
	40710	-54.625*	6	0.001	-78.56	-30.69
40711	15.25	6	0.905	-8.69	39.19	
40801	-14.5	6	0.95	-38.44	9.44	

40806	-42.875*	6	0.001	-66.81	-18.94	
40808	8.375	6	1	-15.56	32.31	
40810	-32.375*	6	0.001	-56.31	-8.44	
40812	-38.875*	6	0.001	-62.81	-14.94	
40815	-54.875*	6	0.001	-78.81	-30.94	
40816	-31.875*	6	0.001	-55.81	-7.94	
40818	-52.000*	6	0.001	-75.94	-28.06	
40823	-11.25	6	0.999	-35.19	12.69	
40901	-7.125	6	1	-31.06	16.81	
40907	-4.125	5.196	1	-24.86	16.61	
40912	14.75	6	0.937	-9.19	38.69	
41009	-2.125	6	1	-26.06	21.81	
41016	8.125	6	1	-15.81	32.06	
41017	14.25	6	0.96	-9.69	38.19	
41108	-31.500*	6	0.001	-55.44	-7.56	
41113	15.125	6	0.914	-8.81	39.06	
41213	15.375	6	0.896	-8.56	39.31	
41215	15.375	6	0.896	-8.56	39.31	
41301	-35.750*	6	0.001	-59.69	-11.81	
41303	-22.25	6	0.122	-46.19	1.69	
41304	-36.250*	6	0.001	-60.19	-12.31	
41305	-20.75	6	0.24	-44.69	3.19	
41401	15.313	5.196	0.634	-5.42	36.04	
41405	15.375	6	0.896	-8.56	39.31	
41703	-26.000*	6	0.014	-49.94	-2.06	
41707	15.375	6	0.896	-8.56	39.31	
41807	15.375	6	0.896	-8.56	39.31	
41901	-22.75	6	0.095	-46.69	1.19	
<b>40210</b>	40118	-2.75	6	1	-26.69	21.19
	40120	-15.25	6	0.905	-39.19	8.69
	40212	-1.125	6	1	-25.06	22.81
	40411	-7.625	6	1	-31.56	16.31
	40503	-26.187*	5.196	0.001	-46.92	-5.46
	40505	-42.125*	6	0.001	-66.06	-18.19
	40510	-37.000*	6	0.001	-60.94	-13.06
	40512	-53.125*	6	0.001	-77.06	-29.19
	40513	-17.125	6	0.71	-41.06	6.81
	40514	-23.937*	5.196	0.005	-44.67	-3.21
	40608	-31.500*	6	0.001	-55.44	-7.56
	40616	-41.125*	6	0.001	-65.06	-17.19
	40702	-7.25	6	1	-31.19	16.69
	40706	-68.000*	6	0.001	-91.94	-44.06
	40710	-69.875*	6	0.001	-93.81	-45.94
	40711	0	6	1	-23.94	23.94
	40801	-29.750*	6	0.001	-53.69	-5.81
	40806	-58.125*	6	0.001	-82.06	-34.19
	40808	-6.875	6	1	-30.81	17.06
	40810	-47.625*	6	0.001	-71.56	-23.69
	40812	-54.125*	6	0.001	-78.06	-30.19
	40815	-70.125*	6	0.001	-94.06	-46.19
	40816	-47.125*	6	0.001	-71.06	-23.19
	40818	-67.250*	6	0.001	-91.19	-43.31
	40823	-26.500*	6	0.01	-50.44	-2.56
	40901	-22.375	6	0.115	-46.31	1.56
	40907	-19.375	5.196	0.115	-40.11	1.36
	40912	-0.5	6	1	-24.44	23.44
	41009	-17.375	6	0.676	-41.31	6.56
	41016	-7.125	6	1	-31.06	16.81
	41017	-1	6	1	-24.94	22.94
	41108	-46.750*	6	0.001	-70.69	-22.81
	41113	-0.125	6	1	-24.06	23.81
	41213	0.125	6	1	-23.81	24.06
	41215	0.125	6	1	-23.81	24.06
	41301	-51.000*	6	0.001	-74.94	-27.06
	41303	-37.500*	6	0.001	-61.44	-13.56
	41304	-51.500*	6	0.001	-75.44	-27.56
	41305	-36.000*	6	0.001	-59.94	-12.06
	41401	0.063	5.196	1	-20.67	20.79
	41405	0.125	6	1	-23.81	24.06
	41703	-41.250*	6	0.001	-65.19	-17.31

	41707	0.125	6	1	-23.81	24.06
	41807	0.125	6	1	-23.81	24.06
	41901	-38.000*	6	0.001	-61.94	-14.06
<b>40212</b>	40118	-1.625	6	1	-25.56	22.31
	40120	-14.125	6	0.965	-38.06	9.81
	40210	1.125	6	1	-22.81	25.06
	40411	-6.5	6	1	-30.44	17.44
	40503	-25.062*	5.196	0.002	-45.79	-4.33
	40505	-41.000*	6	0.001	-64.94	-17.06
	40510	-35.875*	6	0.001	-59.81	-11.94
	40512	-52.000*	6	0.001	-75.94	-28.06
	40513	-16	6	0.841	-39.94	7.94
	40514	-22.812*	5.196	0.011	-43.54	-2.08
	40608	-30.375*	6	0.001	-54.31	-6.44
	40616	-40.000*	6	0.001	-63.94	-16.06
	40702	-6.125	6	1	-30.06	17.81
	40706	-66.875*	6	0.001	-90.81	-42.94
	40710	-68.750*	6	0.001	-92.69	-44.81
	40711	1.125	6	1	-22.81	25.06
	40801	-28.625*	6	0.002	-52.56	-4.69
	40806	-57.000*	6	0.001	-80.94	-33.06
	40808	-5.75	6	1	-29.69	18.19
	40810	-46.500*	6	0.001	-70.44	-22.56
	40812	-53.000*	6	0.001	-76.94	-29.06
	40815	-69.000*	6	0.001	-92.94	-45.06
	40816	-46.000*	6	0.001	-69.94	-22.06
	40818	-66.125*	6	0.001	-90.06	-42.19
	40823	-25.375*	6	0.021	-49.31	-1.44
	40901	-21.25	6	0.194	-45.19	2.69
	40907	-18.25	5.196	0.21	-38.98	2.48
	40912	0.625	6	1	-23.31	24.56
	41009	-16.25	6	0.815	-40.19	7.69
	41016	-6	6	1	-29.94	17.94
	41017	0.125	6	1	-23.81	24.06
	41108	-45.625*	6	0.001	-69.56	-21.69
	41113	1	6	1	-22.94	24.94
	41213	1.25	6	1	-22.69	25.19
	41215	1.25	6	1	-22.69	25.19
	41301	-49.875*	6	0.001	-73.81	-25.94
	41303	-36.375*	6	0.001	-60.31	-12.44
	41304	-50.375*	6	0.001	-74.31	-26.44
	41305	-34.875*	6	0.001	-58.81	-10.94
	41401	1.188	5.196	1	-19.54	21.92
	41405	1.25	6	1	-22.69	25.19
	41703	-40.125*	6	0.001	-64.06	-16.19
	41707	1.25	6	1	-22.69	25.19
	41807	1.25	6	1	-22.69	25.19
	41901	-36.875*	6	0.001	-60.81	-12.94
<b>40411</b>	40118	4.875	6	1	-19.06	28.81
	40120	-7.625	6	1	-31.56	16.31
	40210	7.625	6	1	-16.31	31.56
	40212	6.5	6	1	-17.44	30.44
	40503	-18.562	5.196	0.179	-39.29	2.17
	40505	-34.500*	6	0.001	-58.44	-10.56
	40510	-29.375*	6	0.001	-53.31	-5.44
	40512	-45.500*	6	0.001	-69.44	-21.56
	40513	-9.5	6	1	-33.44	14.44
	40514	-16.312	5.196	0.471	-37.04	4.42
	40608	-23.875	6	0.052	-47.81	0.06
	40616	-33.500*	6	0.001	-57.44	-9.56
	40702	0.375	6	1	-23.56	24.31
	40706	-60.375*	6	0.001	-84.31	-36.44
	40710	-62.250*	6	0.001	-86.19	-38.31
	40711	7.625	6	1	-16.31	31.56
	40801	-22.125	6	0.13	-46.06	1.81
	40806	-50.500*	6	0.001	-74.44	-26.56

\* The mean difference is significant at the 0.05 level.