

Comparative Assessment of Homogeneity Differences in Multi-Temporal NDVI Strata and the Currently Used Agricultural Area Frames in Rwanda

Maurice Mugabowindekwe¹; Gaspard Rwanyiziri^{1,2}.

¹ Centre for Geographic Information Systems and Remote Sensing, College of Science and Technology, University of Rwanda, mugabomaurice@gmail.com, P.O. Box: 3900 Kigali, Rwanda

² Department of Geography and Urban Planning, College of Science and Technology, University of Rwanda, grwanyiziri@gmail.com, P.O. Box: 3900 Kigali, Rwanda

DOI: <http://dx.doi.org/10.4314/sajg.v9i1.7>

Abstract

This study compared two methods used for agricultural statistics generation in Rwanda. The first method is area frame sampling, which is also the currently used method in Rwandan seasonal agricultural surveys; while the second method is the application of remote sensing technique using multi-temporal Normalised Difference Vegetation Index (NDVI) classes to stratify land into homogenous agriculture land classes. The analysis of the methodological flow of Rwanda area frames and the estimated homogeneity in the resulting frames was mainly based on literature review. For the delineation of homogeneous NDVI classes, the study used 10 years data from Moderate Resolution Imaging Spectroradiometer (MODIS) sensor (2004 – 2014). The NDVI data were classified using ISODATA clustering technique, and the focus was put on agriculture-dominated classes, obtained through the intersection with 2010 national land use and land cover data. Analysis of Variance (ANOVA) and Fisher's Least Significant Difference (LSD) statistical methods were applied to investigate significant differences between and within NDVI classes and the currently used Rwanda strata in terms of area coverage of four (4) dominant crops in Rwanda – banana, maize, cassava, and beans. The results of the analysis revealed homogeneity of 85% within NDVI classes, and 69% within the current Rwanda strata, at $p = 0.05$. The NDVI classes were also used to improve the Rwanda strata, and the homogeneity has increased by 5%; reaching 74% after NDVI-based reclassification.

1. Introduction

Continuous growth of the world's population imposes countries to produce more food from limited land resources to ensure that "all people at all times have physical, social and economic access to sufficient, safe and nutritious food which meets their dietary needs and food preferences for an active and healthy life" (Committee on World Food Security, 2009). Agricultural statistics include all information contributing to agricultural development and food security in general; and are key to achieve food security goals (Boyko and Hill, 2009). In addition, agricultural statistics are required to support planning processes of a country; provide information for public policy analysis, debate and

advice; observe agricultural sector performance in the country; monitor and evaluate the impact of policies and programmes; and enlighten decision-making processes (Kiregyera *et al.*, 2007). Moreover, agricultural statistics support the monitoring, evaluation, improvement, and strengthening of early warning systems regarding country's food security (FAO, 2016).

Finally, agricultural statistics contribute to ideas development and decision making of local people, private sector, and importantly, local farmers about their agricultural businesses. Maningas, Villagonzalo, and Macaraig (2004) indicated that agricultural information empowers farmers through having control over their resources and decision-making processes. Nevertheless, usefulness and high quality of the agricultural information depend on three key factors: accuracy – conforming exactly or almost exactly to the ground truth; reliability – showing all details about the reality they represent and errors they contain; and timeliness – availability on time of need (Cotter *et al.*, 2010; Wigton and Bormann, 1978).

The generally used methods to generate agricultural statistics include census, registers, administrative data, and sampling frames (area frame sampling, list frame sampling and multiple frame sampling) (Benedetti *et al.*, 2010; Eurostat, 2015; Kennel, 2008; Väisänen, 2009). However, remote sensing has recently come into use for land stratification to collect and generate agricultural statistics (Boryan *et al.*, Hunt, 2014). This study has focused on two of the above-mentioned methods namely Area Frame Sampling, as currently applied in Rwandan Seasonal Agricultural Surveys (SAS), and Remote Sensing with the use of Normalized Difference Vegetation Index (NDVI) data.

The area frame sampling method is firstly based on construction of *frames* – sets of all agricultural holdings in an area of interest based on their homogeneity, using visual interpretation method (Wigton and Bormann, 1978). Secondly, the *segments* – homogeneous land units enumerable ideally in one day, are drawn from the frames. Finally, representative sample segments are randomly selected for the survey (Kennel, 2008). The method is popular in agricultural surveys because it is believed to be cheaper, reliable, save time, and suitable in case of a large population (Menza *et al.*, 2008). It is used by renowned organizations such as Food and Agriculture Organization of the United Nations (FAO), National Agricultural Statistics Service (NASS) of the United States Department of Agriculture (USDA) among others (Cotter *et al.*, 2010). Although the method is claimed to be accurate with the ability to select representative sample segments for agricultural surveys, it might be an ideal situation different from reality, given the degree of heterogeneity existing in nature (Turner and Gardner, 1994). Remote sensing is also currently used for land stratification to generate strata for agricultural surveys (Carfagna and Gallego, 2005).

The aim of this research was to apply multi-temporal NDVI data from MODIS, to stratify land in Rwanda and generate agricultural statistics for the four (4) main crops in Rwandan agricultural season A (September – February of the following year) for 2015 – 2016 agricultural year, and make a comparative assessment of homogeneity differences in both currently used area frames and the NDVI strata.

2. Materials and methods

2.1. Study area

The study area for the present research was not based on administrative boundaries. The NDVI stratification determined the representative areas for field data collection. The study area was composed of four NDVI Classes: NDVI class 24, NDVI class 54, NDVI class 70 and NDVI class 82 (Figure 1). Classes were spatially distributed into 7 administrative districts. Administratively, NDVI class 24 was mainly located in Bugesera district (Eastern Province); class 54 in Nyaruguru and Nyamagabe districts (Southern province); class 70 in Nyabihu district (Western Province) and Musanze district (Northern Province); and class 84 had a part in Musanze and Burera districts (Northern province), and another part in Nyagatare district (Eastern Province) as presented in Figure 1 below.

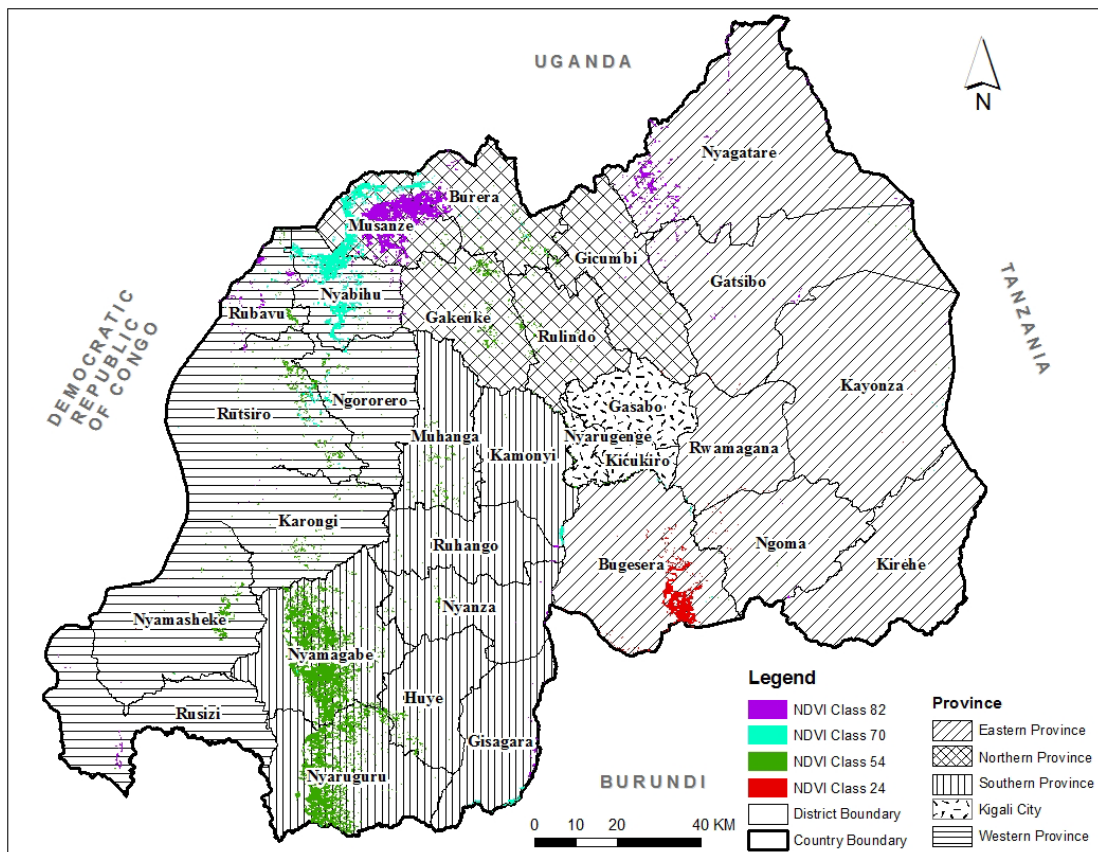


Figure 1. Selected Sample NDVI Classes for the Study Area

Class 24 covered an area of about 6,485 ha; class 54 covered 57,031 ha; class 70 covered 19,583 ha; and class 82 covered an area of about 20,769 ha, making a total area of 103,868 ha which is equal to 9.3% of entire cropland area in Rwanda, as of 2010.

Agro-ecologically, as illustrated by Figure 2, class 24 was majorly situated in Eastern Savannah of Rwanda: a region with an average altitude of 1,400 m above mean sea level (AMSL), 850 mm of average annual rainfall, mean annual temperature of around 21°C, with old infertile soils with a

variable texture. Class 54 was majorly located in both Congo-Nile divide (area with an average altitude of about 2,100 m AMSL, 1,600 mm of mean annual rainfall, mean annual temperature of around 19°C, with soil dominated by humic soils at medium altitude), and central plateaus (area with average altitude of about 1,700 m AMSL, 1,200 mm of average annual rainfall, mean annual temperature of around 17°C, with soil dominated by ultisols derived from volcanic materials). Class 70 was located majorly in volcanic summits and high plains (area with an average altitude of about 2,200 m AMSL, 1,500 mm of average annual rainfall, mean annual temperature of about 17°C, with soil dominated by ultisols derived from volcanic materials), and Congo-Nile divide. Finally, class 82 was majorly located in both volcanic summits and high plains and Eastern Savannah.

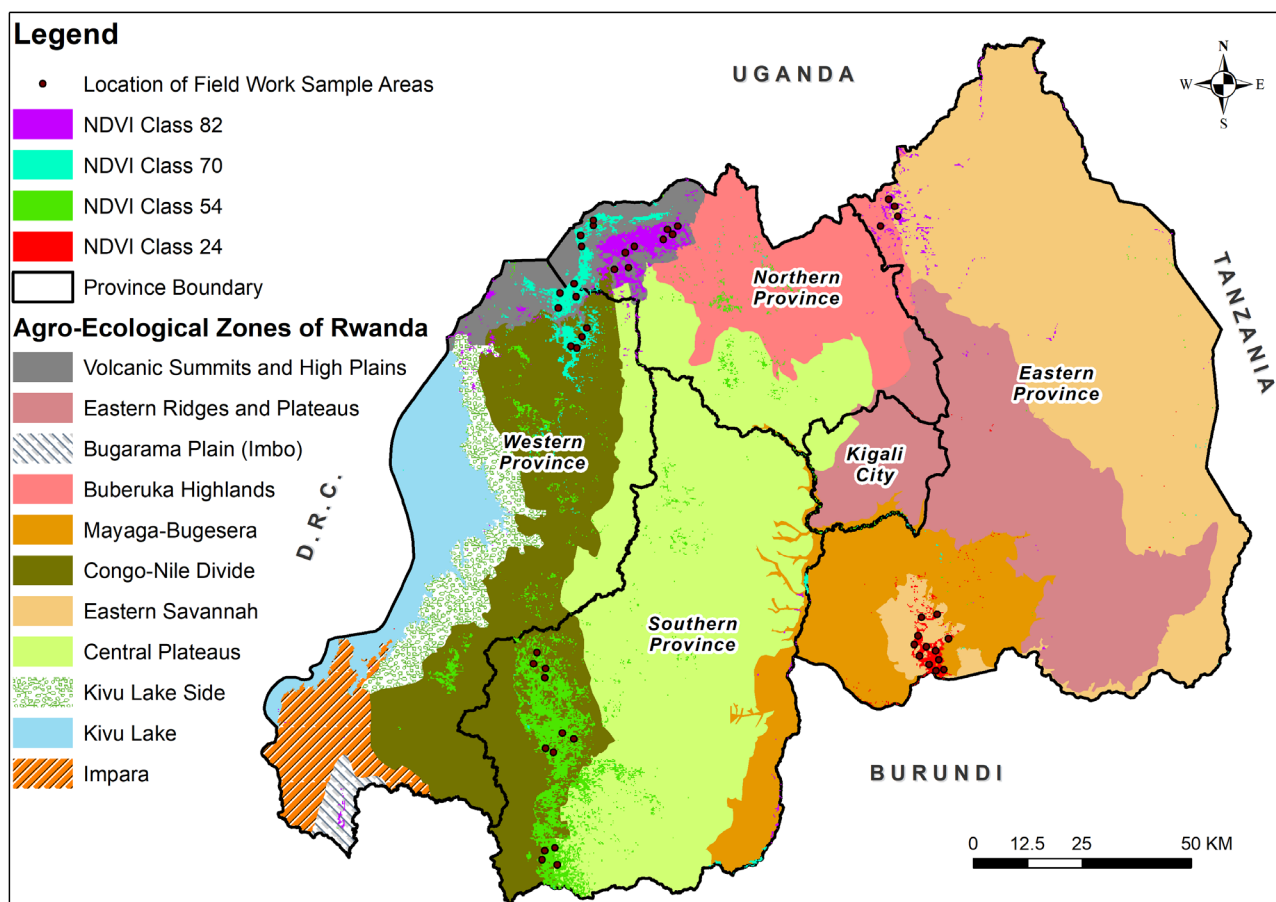


Figure 2. Location of Sample NDVI Classes and Field Work Areas in Rwandan Agro-ecological Zones

2.2. NDVI data

Different NDVI products such as those from MODIS, AVHRR, MERIS sensors, etc., have been used with success for various studies; either for natural phenomena studies or agro-ecological studies (Barbosa *et al.*, 2006; Benedetti and Rossini, 1993; Bolton and Friedl, 2013; Funk and Budde, 2009). The rationale for selecting NDVI was that it is one of the radiometric measures of the amount of greenness on the land according to the photosynthetic activity present on that land (NASA, 2015; Tucker and Sellers, 1986). This attribute justifies the ability of NDVI to group together areas from

which vegetation cover appears similar in terms of density and content, to maximise the homogeneity within the same stratum.

This study applied NDVI data from MODIS sensor onboard Terra satellite with 250 m spatial resolution (MOD13Q1 products). The main reasons were – quick revisit time, a reasonable spatial resolution given the extent of the study area, and wide swath covering a large area. Data are produced as 16 days maximum value composites to reduce effects of clouds, directional reflectance and off-nadir viewing effects, sun angle and shadow effects, and aerosol and water vapour effects (Holben, 1986; King *et al.*, 2003). In order to ensure continuous and long-term homogeneity consideration in sites within the study area, the data's temporal coverage considered in the present study is 10 years (2004 – 2014).

2.3. NDVI data pre-processing and processing

In total, 221 MODIS-NDVI images (MOD13Q1 products) from 2004 to 2014 were acquired from MODIS data pool using MODIS Reprojection Tool Web interface (MRTWeb). They were rescaled from original values [ranging from -3,000 to 10,000 and stored under 16-bit signed integer (LP DAAC, 2014)] to the normal pixel value range (0 to 255 for pixels stored under 8 bits), and later, normalised to the normal NDVI range from -1 to 1. The rescale applied the following equation (de Bie, 2015):

$$\text{Rescaled NDVI} = \text{Original NDVI} * 0.02125 + 43$$

Equation 1: NDVI rescaling

Where *Rescaled NDVI* refers to the obtained NDVI values after the rescaling process (ranging between 0 and 255, and later, normalised to the normal NDVI range from -1 to 1). *Original NDVI* refers to the original NDVI values as in originally acquired MOD13Q1 images (ranging between -3,000 and 10,000). Obtained images were stacked and the layer stack was cleaned and smoothed using Savitzky – Golay smoothing and differentiation filter, in order to fit the data points optimally to a polynomial in the least square logic, to eliminate noise in the data and fit them with optimal data (Luo *et al.*, 2005).

2.4. Land stratification based on MODIS – NDVI and representative sampling

The unsupervised classification was performed on the 221 MOD13Q1 images in order to determine similar classes. These images were following the general phenological characteristics of crops in Rwanda: where ploughing, planting and germination occur in September, growing of most of the crops in October and November, and the maturity stage and harvesting in December and January, for Season A. Seasons B, goes with ploughing, planting and germination in February, growing of most of the crops happen in March and April, and the maturity stage and harvesting happen in May and June. This can be observed later in Figure 6. The classification applied the Iterative Self-Organizing Data Analysis Technique (ISODATA) (Ball and Hall, 1965; in

Memarsadeghi., 2007). Minimum number of classes taken was 10 and the maximum was 100, with 50 as maximum number of iterations, and 1.0 as convergence threshold (Ali et al., 2012). Separability analysis using average and minimum separability was used to find optimal number of classes, which was found to be 95 for the country (Figure 5). The classes were overlaid with recent Rwanda land use data (2010 data), and it was revealed that only 24 classes contain more than 50% of agriculture. In the end, the sampling was drawn from the 24 agricultural NDVI classes. The 2010 Rwanda land use data were produced by former Rwanda Natural Resources Authority (RNRA), currently Rwanda Land Management and Use Authority (RLMUA) in partnership with Regional Centre for Mapping of Resource for Development (RCMRD) using both Landsat data, image classification and intensive field survey for validation. Moreover, the data considered the “Kagera Agro-Ecosystems” Land use map of Rwanda prepared during a workshop held in Huye district, Rwanda, at the Centre for GIS and Remote Sensing of University of Rwanda (CGIS – UR), from the 22nd of November to the 3rd of December 2010 in the framework of the Kagera TAMP (Transboundary Agro-ecosystem Management Project for the Kagera river basin) project, under implementation of FAO and financial support from Global Environmental Facility (GEF) (FAO, 2010). The land use data contained various land-use types including cropland, forestland, wetland, grassland, settlement and other land.

2.5. Random-representative clustered sampling

The size of a sample area was equal to the pixel size of the used MODIS NDVI data; which was originally 231.66 * 231.66 (53,666 m² ~ 5.37 ha) after the data acquisition. After data conversion and projection to match with Rwandan land use data, the final pixel size became 231.92 * 230.37 m ~ 5.34 ha. Similarly, conversions and geometric transformations resulted into a pixel shift of 1.69 m to the right (for the upper corner of NDVI data pixels), and 2.51 m to the right too (for the lower corner). However, the shift could not meaningfully affect the analysis, given the size of the pixel and aim of the study, considering the avoided possibility of half a pixel shift.

The available fieldwork time prompted the selection of 12 sample sites per sample NDVI class. Nonetheless, one site was excluded after visiting the area and realized that there was no vegetation cover due to the recent conversion from agriculture into industrial zone construction (Figure 4a and 4b). Henceforth, the final number of sample areas was 47. Sample areas were selected randomly given that they fulfil three conditions. First, being located in the innermost part of a class, to avoid influences from outside. Second, being located close to the road, for accessibility. Finally, the spatial distribution of sample sites had to follow a pattern of four (4) sample sites (first cluster) in the North or West of a class, second cluster in the Centre, and last cluster in the South or East of the class (Figure 2).

2.6. Current area frame sampling in Rwanda

The assessment of the current area frame sampling method, as applied in Rwanda, was carried out mainly based on literature review. The used documents explain not only basic principles, procedures and rationale for using the technique in an agricultural survey (Cotter et al., 2010; Wigton and

Bormann, 1978) but also how the technique was specifically applied in 2012 – 2013, 2013 – 2014 and 2014 – 2015 Rwanda Seasonal Agricultural Surveys (NISR, 2013, 2015c, 2015b, 2015a). Moreover, the used documents clarify that area frames for the surveys in Rwanda are still constructed using 2008 aerial photographs, by the support of visual interpretation technique, to delineate “homogeneous” strata (area frames) based on crops intensity. From the defined area frames, segments are delineated by the use of physical features such as roads, paths, rivers, etc.

2.7. Data collection

Land cover types and their area coverage were collected in the field for two months (October – November 2015). They were collected using a printed field map with the 2008 high-resolution aerial photographs, and the field observation (with recording of geographical coordinates using GPS), and unstructured interviews with local farmers found in their agricultural field during the fieldwork. The field data were entered into computer system through digitisation process to get final interpreted data as shown in the following maps (Figure 3a and 3b).

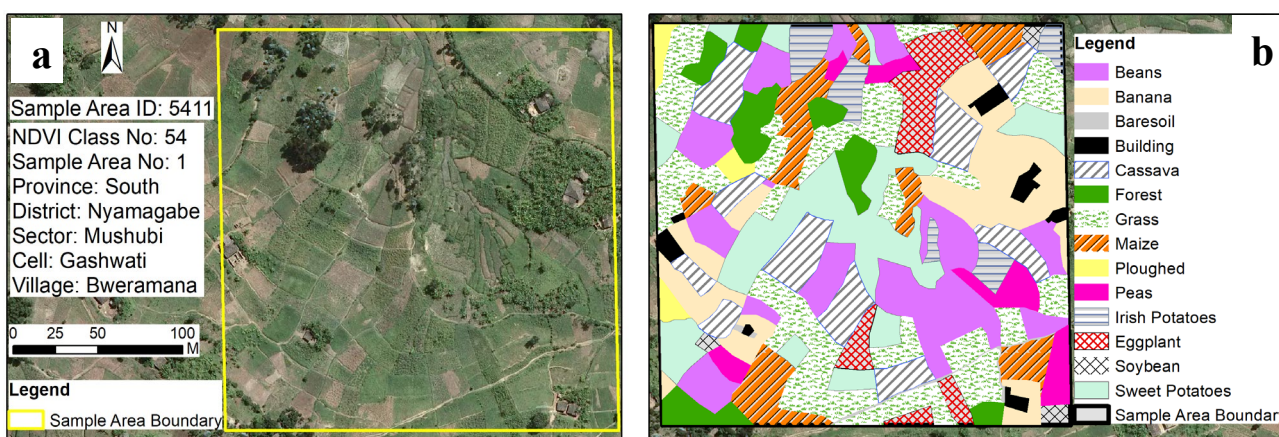


Figure 3. Example of a Field Map before Data Collection (a) and Interpretation after Field Work (b)

After data entry, the area of each land cover type was calculated and presented in a table for further analysis. Table 1 is an example of land cover area data from the sample site 5411 in Figure 3.

Table 1. Land Cover Area Data Collected from the Sample Site 5411 in Figure 3

Cover	Area (ha)	Area (%)	Cover	Area (ha)	Area (%)
Banana	0.61	11.49	Grass	1.03	19.35
Beans	0.77	14.47	Irish potatoes	0.20	3.69
Bare Soil	0.02	0.37	Maize	0.43	8.12
Building	0.08	1.52	Peas	0.19	3.54
Cassava	0.59	11.01	Sweet potatoes	0.72	13.32
Eggplant	0.27	5.07	Ploughed	0.07	1.36
Forest	0.32	5.94	Soybean	0.04	0.75
Total				5.34	100

2.8. Data analysis

Data analysis took into account area coverage data (ha) of the four main crops of Rwandan agricultural season A: banana, maize, beans and cassava. Analyses were carried out in four stages for NDVI strata, and two stages for Rwandan Strata. Firstly, one-way ANOVA was applied to find out whether the NDVI classes were significantly different between them, to be considered effective as separable classes. Secondly, Fisher's Least Square Difference (LSD) was carried out for the classes' pairwise comparison, to identify specific differences between the classes. Thirdly, one-way ANOVA was carried out for clusters within the same NDVI class to evaluate the level of possible significant differences between clusters within the same class, to ensure that every single NDVI class is homogeneous (if no significant differences between its clusters), or otherwise heterogeneous. Fourthly, only in case of significant differences within a class, Fisher's LSD was applied to identify specific differences among clusters pairs of the class, leading to the quantification of the level of homogeneity and heterogeneity of the class.

However, for Rwandan Strata, most of the sample sites (40 out of 47) were located in Rwandan stratum 1. Therefore, the analysis was limited to the Rwandan stratum 1, and two first stages of analysis, given the fact that there were not enough data to make statistical analysis for the rest of the strata.

3. Results and discussion

3.1. Land stratification with the currently used area frame sampling in Rwanda

Currently, area frames for Rwanda seasonal agricultural surveys are designed using the 2008 aerial photographs. By the use of visual interpretation technique, interpreters consider areas with similar crops' intensity and delineate them as homogeneous strata. Table 2 presents the designed strata for the 2012 – 2013 seasonal agricultural survey.

Table 2. Rwanda Seasonal Agricultural Survey Strata

Strata	Description	Land Area (ha)
1	Intensive hillside cropland (50-100% cultivated)	1,535,000
2	Intensive marshland cropland (50-100% cultivated)	55,100
3	Extensive cropland (15-50% cultivated)	192,800
4	Non-cropland (0-15% cultivated)	73,900
5	Cities and towns (0-15% cultivated)	47,700
6	Water	130,200
7	National parks (defined by political boundaries)	219,000
8	Marshlands, riverbeds with potential for rice (0-15% cultivated)	79,200
9	Forest	172,200
10	Tea plantation	23,200

Source: NISR (2015b)

Three strata dominated by agriculture (Rwanda Stratum 1 – intensive hillside cropland; Rwanda Stratum 2 – intensive marshland cropland; and Rwanda Stratum 3 – extensive cropland) were used for further sampling design and survey. Rwanda stratum 1 covered 86% of the entire survey area, and stratum 2 and stratum 3 covered only 3% and 11% of the total area respectively.

The three survey strata were divided into Primary Sampling Units (PSUs) using physical boundaries (roads, paths, rivers, etc.), from which representative sample PSUs were randomly selected using Probability Proportional to Size (PPS) technique. Per sample PSU, Secondary Sampling Units (SSUs or segments) were constructed, and one segment was randomly selected as a representative for the PSU. The size of one segment was about 20 hectares, except for stratum 3 where a segment was about 50 ha in size. For instance, a PSU of 225 ha was divided into 11 segments of approximately 20 ha each, and one segment was randomly selected for the survey to represent the PSU (NISR, 2013). The total number of sample segments was 327, where 295 segments were for stratum 1; 14 in stratum 2; and 18 in stratum 3.

Rwanda strata data types and level of up-to-date

The currently used area frame sampling in Rwanda uses the 2008 aerial photographs for the area frames design. The photographs have been in use for seasonal agricultural surveys (SAS) since 2012 till up to now (2020). The use of outdated spatial data for agricultural surveys may result in biased sample strata selection which may lead to erroneous data representativeness and error propagation into overall national agricultural aggregates. This is supported by the argument of Roser (2015), that agriculture is among the most rapidly changing land uses, consequently, it is necessary to rely on up-to-date data for agricultural surveys.

Rwanda strata data's spatial and temporal richness

Aerial photographs used to define the currently used area frames in Rwanda are of high spatial resolution (0.25 m * 0.25 m) (MINIRENA and RNRA, 2009). However, the photographs are one – time (2008) data in terms of temporal coverage. Therefore, they have a weakness in encountering spatial-temporal land cover and land-use changes that took place after the aerial survey.

Rwanda strata and sampling process with consideration of natural phenomena

Land stratification for the currently used area frames in Rwanda is done through visual interpretation technique, which might introduce errors. In fact, the stratification depends on personal perception and professional experience of the interpreter (Baks *et al.*, 2013). This is not only increasing the level of bias in stratification and sampling design but also it might ignore important natural phenomena, such as possible abrupt changes in the natural system given the rapid changes in agricultural land cover. This is exemplified by the area in figure 4b where agricultural activities were cleared 2 months prior to the fieldwork. The land clearance was for the favour of new industrial zone construction. However, these changes do not appear in the 2008 aerial photograph (Figure 4a) and yet, the area had been classified as agriculture for different seasonal agricultural surveys since 2012.

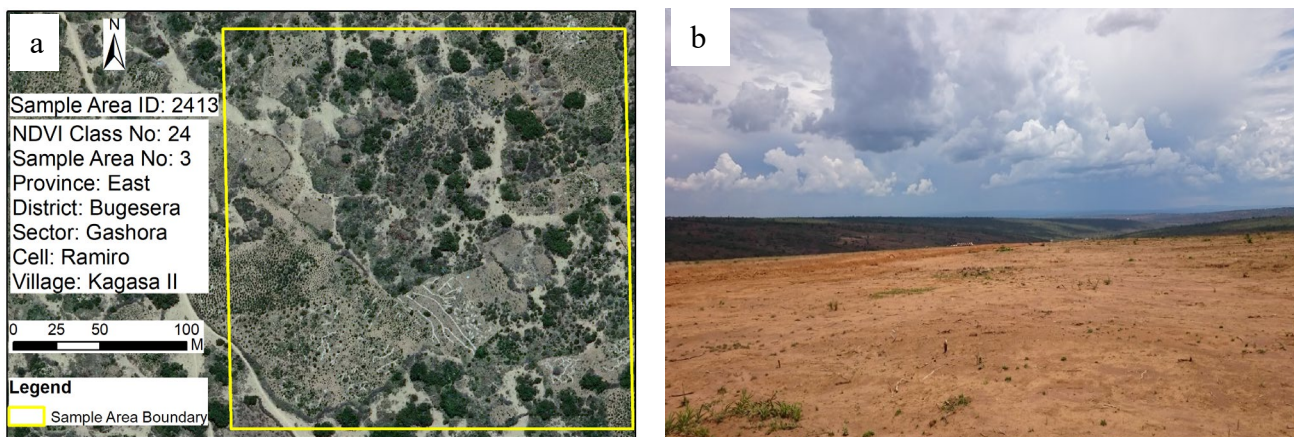


Figure 4. 3rd Surveyed Sample Site as Appeared in 2008 (a) and in 2015 after Site Clearing (b)

Rwanda strata validity period

Cotter et al. (2010) argued that the currently designed area frames can be used for 15 to 20 years without a need to be replaced. The currently designed area frames in Rwanda are valid for 5 years as informed by staff in charge of national agricultural surveys, at the National Institute of Statistics of Rwanda. Nevertheless, within this period, many changes occur and they might introduce errors in the subsequent surveys.

3.2. Land stratification using multi-temporal NDVI data

Separability analysis of NDVI classes using average and minimum separability values of the classification results from 10 to 100, revealed that an optimal number of classes to distinguish heterogeneous areas in Rwanda from 2004 to 2014 is 95 (Figure 5).

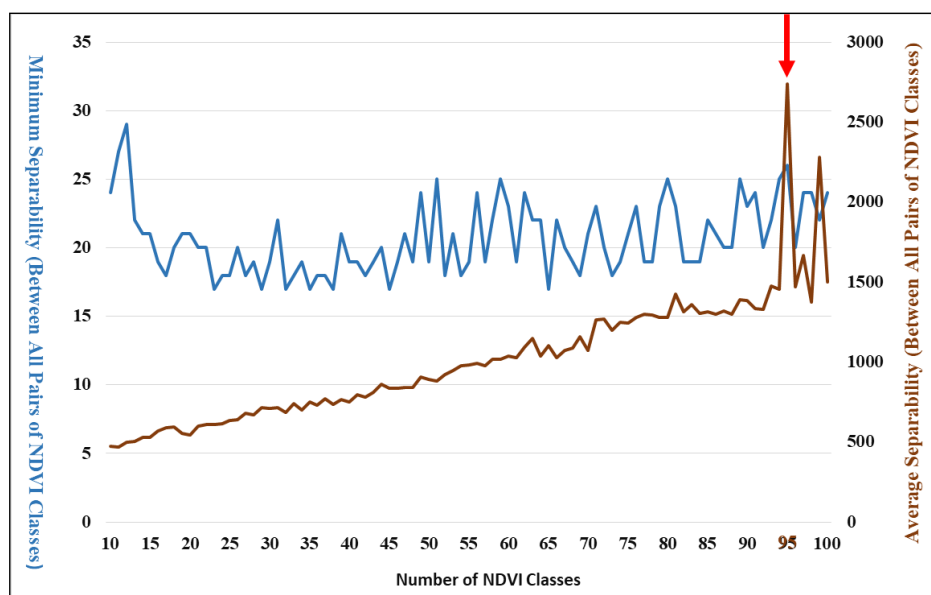


Figure 5. Separability Analysis of NDVI Classes

The 95 classes were overlaid with the 2010 Rwanda land use data, and 24 classes were found to be dominated by agriculture ($\geq 50\%$). For 10 years, a class had had similar behaviour in terms of vegetation intensity, but with some changes in some classes. For representative sampling and time available for the research, four (4) sample NDVI classes were purposively selected given that all classes from those with low vegetation intensity to those with high vegetation intensity throughout agricultural seasons are represented (Figure 6).

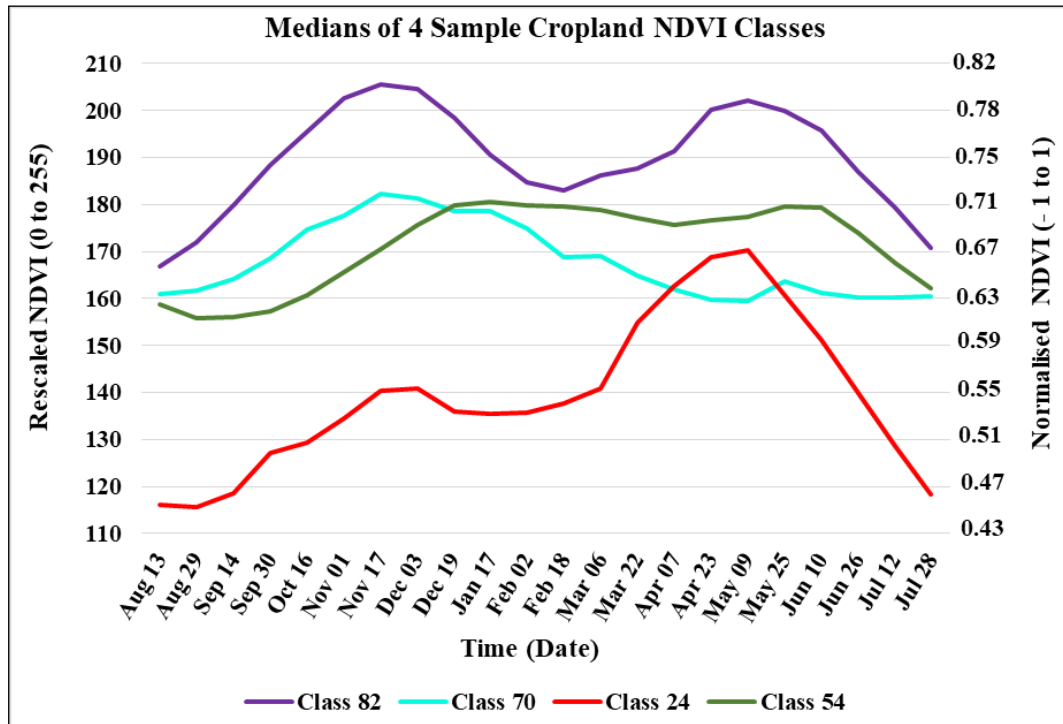


Figure 6. Average Annual Medians of the Four Sample Agricultural NDVI Classes

Distinctive temporal behaviour of the sample NDVI classes led to the differences in spatial distribution as well and; covering parts of 7 districts out of 30 districts, from South – West to North – East (Figure 1) in different agro-ecological zones of the country (Figure 2).

NDVI data types and level of up-to-date

NDVI classes are defined based on the amount of greenness present in a specific area over time (NASA, 2015). This indicates that, no matter how the vegetation cover of an area is rapidly changing, if it changes with a similar pattern throughout a year and repeatedly over time, the area should be considered as the same class. This makes it an important tool in agricultural surveys, which deals with rapidly changing land use. Furthermore, most of the NDVI data are regularly up-to-date. For instance, MODIS data used by the present study are available every two days since 1999 (NASA, 2002). Other data include NOAA-HVRR data which are available every day since 1981 (Tucker, 2009), but with coarser spatial resolution, thus suitable for surveys over a big sized land such as of a continental scale. Additionally, most of the multi-temporal NDVI data are open and free of charge.

NDVI data's spatial and temporal richness

NDVI data applied by the present study are not very rich in the spatial dimension, as they aggregate a 250 m² into one unit with the same reflectance value “a pixel” (Liew, 2001). However, they are very rich in the temporal dimension. Moreover, spatial richness was claimed back through integrating the data with recent land-use data (RNRA, 2010) with a high spatial richness to allow the knowledge of each land cover contained into a single pixel area and its coverage (in ha and %).

NDVI stratification process and sampling with consideration of natural phenomena

NDVI data consider events occurring in nature according to similarities in areas' vegetation photosynthetic activity (Tucker and Sellers, 1986). Nonetheless, the area recently converted for industrial zone construction (Figure 4a and 4b) was classified as agriculture because the present study did not include NDVI image data of 2015; a year when the changes took place. However, if fieldwork would take place the following year (2016), the 2015 image data would have been included, for the recent changes to be considered as well for the new stratification. Hence, representativeness is improved by the use of up-to-date data. Additionally, the land classification process is done with the use of the computer system after the user's instructions, through unsupervised classification, which reduces the bias from the interpreter and promotes replicability of the process.

NDVI strata time frame validity period

NDVI-based land stratification should be updated every agricultural year, thanks to the regular availability of up-to-date data. The regular data updates ensure that recent spatial-temporal heterogeneity is incorporated into new land stratification for the subsequent agricultural surveys. This would regularly improve homogeneity within the strata of new surveys.

Land cover types and area coverage in Rwandan strata

The land cover data collected from the field per sample NDVI class were overlaid with the currently used area frames. It was found that among the 47 total sample sites, 40 were fully located in Rwanda stratum 1 “Intensive hillside cropland”; 1 sample site fully in Rwanda stratum 5 “cities and towns”, 2 sample sites fully in Rwanda stratum 9 “forest”, 2 sample sites in Rwanda stratum 10 “tea plantation”, 1 sample site partly in Rwanda stratum 1 and partly in Rwanda stratum 9, and 1 sample site partly in Rwanda stratum 1 and partly in Rwanda stratum 5 (Figure 7). In the end, only the data for sample sites in Rwanda stratum 1 were enough to make further statistical analyses.

The land cover types identified include main crops of agricultural season A (banana, maize, beans, and cassava), and other different crops and cover types. However, looking at the presence and intensity of crops and land cover types, the cover types found in Rwandan Strata do not present similar patterns in intensity, as it would be expected for cover types from the same stratum which is assumed “homogeneous”. In addition, there are crops and cover types which were dominantly found in some sample sites but were not found in other sites from the same stratum. This might imply that the considered “same” stratum might be presenting a too low level of homogeneity to be considered entirely homogeneous.

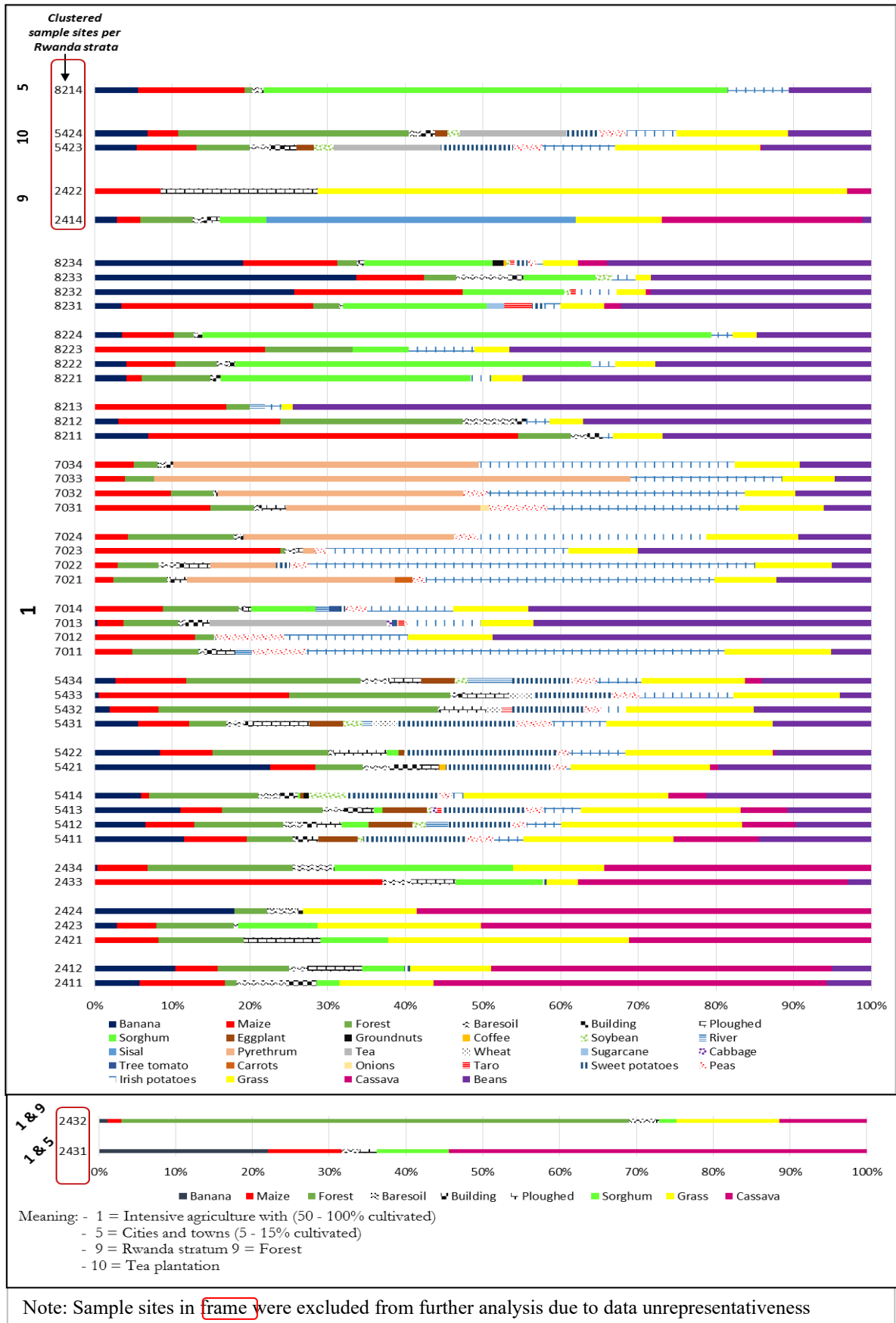


Figure 7: Land Cover Types in Rwanda Strata

Land cover and area coverage in NDVI classes

There has been identified a remarkable patterns of land cover intensity according to their respective NDVI classes and clusters. Looking at dominant land cover types, for instance, cassava dominated class 24, but it was hardly present in other classes. The grass was abundant in class 24 and 54, but little in class 70 and very little in 82. Sorghum was found in all classes, but dominant in class 82, little in class 24, very little in class 54 and only in one sample site in class 70.

Beans dominated class 82 and 70 and was more dominant in cluster 1 of class 70, where it corresponded to local farmers' information that the area is part of the region selected by policy-makers to be "a beans model plots area". Banana (together with maize and beans) was dominant in class 82, especially in cluster 3. This might be due to the effect of crop intensification policy (Cantore, 2013; Kathiresan, 2012), as the interviewed farmers informed that their region is one of the regions allowed to grow banana under the ongoing crop regionalisation policy.

Maize had almost similar proportions in all NDVI classes, except in class 82, where it was among the abundant crops (in addition to beans and banana) compared to other classes. Similarly, the forest was present in all NDVI classes but dominant in class 24 and 54. Although from this, it would be argued that forest segregation might be difficult to achieve in different areas of Rwanda using coarse resolution NDVI data, as the focus of this study was mainly on crops, it would be gladly recommended for future similar studies to try the relatively high (or moderate) resolution data such as Landsat OLI and/or Sentinel 2 NDVI data, for the forest cover differentiation in Rwandan heterogeneous landscapes.

Furthermore, some of the crops and other land cover types such as onions, carrots, sisal, cabbage, sugarcane, wheat, tea, coffee, groundnuts, peas, eggplants, tree tomatoes, taro, and soybean were not frequent and sometimes found only in one sample site, hence, these very dominated crops can also better be studied with high-resolution NDVI data, given their limited spatial coverage in most of the Rwandan landscapes.

Statistical analysis of variance of crops cover area in Rwandan strata

The following Table 3 presents area size of main season A crops cover per sample site, as collected from the field.

Table 3. Rwandan Season A Main Crops Cover Area Size (in ha) per NDVI Sample Site

Crops	Banana				Maize				Beans				Cassava			
	Class 24	Class 54	Class 70	Class 82	Class 24	Class 54	Class 70	Class 82	Class 24	Class 54	Class 70	Class 82	Class 24	Class 54	Class 70	Class 82
11	0.31	0.61	0	0.37	0.58	0.43	0.26	2.54	0.31	0.77	0.28	1.44	2.70	0.59	0	0
12	0.56	0.35	0	0.16	0.29	0.34	0.69	1.11	0.27	0.52	2.60	1.98	2.35	0.37	0	0
13	-	0.59	0.02	0	-	0.28	0.18	0.90	-	0.58	2.33	3.98	-	0.32	0	0
14	0.15	0.32	0	0.30	0.16	0.06	0.47	0.73	0.06	1.13	2.36	0.57	1.38	0.26	0	0
21	0	1.21	0	0.22	0.44	0.31	0.13	0.10	0	1.06	0.65	2.40	1.67	0.05	0	0
22	0	0.45	0	0.22	0.46	0.36	0.16	0.33	0	0.66	0.27	1.49	0.17	0.01	0	0
23	0.15	0.29	0	0	0.27	0.41	1.28	1.17	0	0.75	1.60	2.49	2.68	0.01	0	0
24	0.96	0.36	0	0.19	0	0.21	0.23	0.35	0	0.57	0.50	0.79	3.12	0	0	0
31	1.18	0.30	0	0.18	0.51	0.35	0.79	1.32	0	0.68	0.33	1.72	2.91	0	0	0.12
32	0.06	0.11	0	1.37	0.10	0.33	0.53	1.16	0	0.81	0.52	1.51	0.61	0	0	0.04
33	0	0.03	0	1.80	1.98	1.31	0.21	0.47	0.16	0.22	0.25	1.51	1.85	0	0	0
34	0.02	0.15	0	1.02	0.35	0.49	0.27	0.65	0	0.75	0.49	1.81	1.84	0.12	0	0.20

Results of ANOVA showed that there were 75% of significant differences (heterogeneity), and 25% of non-significant differences (homogeneity) between clusters within Rwanda stratum 1 in terms of the main crops at 95% confidence level (p was $4e-4 < \alpha$ (0.05) in terms of Banana cover; and $0.07 > \alpha$ (0.05) in terms of Maize cover; and $2e-4 < \alpha$ (0.05) in terms of Beans cover; and $1e-16 < \alpha$ (0.05) in terms of Cassava cover). However, the pairwise comparison using Fisher’s LSD indicated that, overall, Rwanda stratum 1 was 31% heterogeneous (specific significant differences) and 69% homogeneous (specific non-significant differences) at 95% confidence level.

Statistical analysis of variance of crops cover area in NDVI classes

The ANOVA results revealed that NDVI classes were significantly different in NDVI classes (p -value was $0.02 < \alpha$ (0.05) in terms of Banana; $0.04 < \alpha$ (0.05) in terms of Maize; $2.02e-6 < \alpha$ (0.05) in terms of Beans; and $2.25e-13 < \alpha$ (0.05) in terms of Cassava), at 95% confidence level. Therefore, the classes were significantly separable. To investigate specific significant differences within the classes, Fisher’s LSD revealed that among total pairs of clusters within NDVI classes; there were 15% significant differences (heterogeneous), and 85% no significant differences (homogeneous), at 95% confidence level. All in all, NDVI classes presented higher homogeneity in terms of the four main crops, compared to the currently used Rwanda Stratum 1.

Use of NDVI stratification to improve homogeneity within Rwanda stratum 1

To analyse the possibility to improve currently used Rwanda strata, the misclassified sample sites within Rwanda strata were correctly classified according to the ground truth collected in accordance with NDVI classification. This resulted in 4 sample sites being reclassified as part of Rwanda stratum 1. ANOVA results for the reclassified Rwanda stratum 1 showed that, then, significant differences were lowered to 26%, and non-significant differences increased to 74%, at 95% confidence level.

Thus, homogeneity in Rwandan stratum 1 was improved by 5% thanks to improved classification based on NDVI.

4. Conclusion

The present study carried out a comparative assessment of homogeneity differences in both currently used area frame sampling in Rwandan Seasonal Agricultural Surveys, and multi-temporal NDVI stratification. The land use and land cover data collected from the field were used to assess homogeneity differences in both Rwanda stratum 1 and NDVI strata, in terms of Rwanda season A main crops (banana, maize, beans, and cassava). However, the assessment of other aspects of current Rwanda strata such as techniques used to stratify land, data types, etc., was mainly based on literature review.

NDVI stratification resulted in strata different in both temporal behaviour and spatial distribution. Furthermore, sample NDVI classes were found to be significantly different in terms of Rwandan agricultural season A main crops at 95% confidence level, which confirmed their separability for effective representativeness in an agricultural survey.

Land stratification using NDVI and the currently used area frame sampling were found to have differences and similarities, but this study found NDVI stratification to be more reliable for improved homogeneity within classes while delineating homogeneous strata for agricultural survey. The most important difference lied in the homogeneity encountered in classes. NDVI classes were found to encounter a higher level of homogeneity compared to the current Rwanda strata, in terms of the main crops at 95% confidence level. ANOVA and Fisher's LSD results revealed 85% and 69% of homogeneity within NDVI classes and Rwanda stratum 1 respectively. Moreover, NDVI stratification was used to improve the homogeneity within the Rwandan Stratum 1, and this resulted in a 5% homogeneity improvement for the latter, reaching 74%.

Comparing aspects such as level of updating the data used for stratification, types and availability of data, and time of land stratification validity, it was found that land stratification for agricultural surveys using NDVI data has more potentials for accurate agricultural statistics and reduction of error propagation, compared to the currently used area frame sampling method.

5. Acknowledgements

The authors acknowledge the priceless effort and the contribution of various people and organizations for the success of this research. We are deeply thankful to the former Netherlands Fellowship Program (NFP), currently Orange Knowledge Programme (OKP), for financial support. Furthermore, we are very grateful to Drs. Edward Westinga and Dr. Ir. C.A.J.M. (Kees) de Bie from the department of Natural Resources at ITC – Faculty of Geo-Information Science and Earth Observation of University of Twente, for tireless effort in guidance and advice to achieve the present

scientific output. Moreover, the authors gratefully acknowledge the contribution of the National Institute of Statistics of Rwanda, local government authorities and local population of Nyaruguru, Nyamagabe, Bugesera, Nyabihu, Musanze, Burera, and Nyagatare districts in Rwanda, for authorising and facilitating field data collection. Finally, authors would like to acknowledge the great effort of the anonymous reviewers who took their time and provided very insightful, thoughtful and instructive comments which surely helped to leverage the quality of this research output.

6. References

- Ali, A., de Bie, C. A. J. M., Scarrott, R. G., Ha, N. T. T., & Skidmore, A. K. (2012). Comparative Performance Analysis of a Hyper-Temporal NDVI Analysis Approach and a Landscape-Ecological Mapping Approach. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 1–7(1), 105–110. <https://doi.org/10.5194/isprsannals-I-7-105-2012>
- Baks, W., Janssen, L., Schetselaar, K., & Tolpekin, V. (2013). Image Analysis. In V. Tolpekin & A. Stein (Eds.), *The Core of GIScience: A Process Based Approach* (ITC Educat). Enschede: The International institute for Geo-Information Science and Earth Observation (ITC), University of Twente.
- Ball, G. H., & Hall, D. J. (1965). *ISODATA, A Novel Method of Data Analysis and Pattern Classification*. Menlo Park, California.
- Barbosa, H. A., Huete, A. R., & Baethgen, W. E. (2006). A 20-Year Study of NDVI Variability over the Northeast Region of Brazil. *Journal of Arid Environments*, 67(2), 288–307. <https://doi.org/10.1016/j.jaridenv.2006.02.022>
- Benedetti, R., Bee, M., Espa, G., & Piersimoni, F. (Eds.). (2010). *Agricultural Survey Methods*. Chichester, UK: John Wiley & Sons, Ltd. <https://doi.org/10.1002/9780470665480>
- Benedetti, R., & Rossini, P. (1993). On the Use of NDVI profiles as a Tool for Agricultural Statistics: The Case Study of Wheat Yield Estimate and Forecast in Emilia Romagna. *Remote Sensing of Environment*, 45(3), 311–326. [https://doi.org/10.1016/0034-4257\(93\)90113-C](https://doi.org/10.1016/0034-4257(93)90113-C)
- Bolton, D. K., & Friedl, M. A. (2013). Forecasting Crop Yield using Remotely Sensed Vegetation Indices and Crop Phenology Metrics. *Agricultural and Forest Meteorology*, 173, 74–84. <https://doi.org/10.1016/j.agrformet.2013.01.007>
- Boryan, C. G., Yang, Z., Di, L., & Hunt, K. (2014). A New Automatic Stratification Method for U.S . Agricultural Area Sampling Frame Construction Based on the Cropland Data Layer. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(11), 4317–4327.
- Boyko, E., & Hill, C. (2009). *An Introduction to Agriculture Statistics*. Maputo. Retrieved from http://siteresources.worldbank.org/SCBEXTERNAL/Resources/Introduction_to_Agriculture_Statistics.pdf
- Cantore, N. (2013). The Crop Intensification Program in Rwanda: a Sustainability Analysis. Kigali: Overseas Development Institute. Retrieved from <https://www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/7094.pdf>
- Carfagna, E., & Gallego, J. F. (2005). Using Remote Sensing for Agricultural Statistics. *International Statistical Review*, 73(3), 389–404. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1111/j.1751-5823.2005.tb00155.x/abstract>
- Committee on World Food Security. (2009). Declaration of The World Summit on Food Security. Rome: Food and Agriculture Organisation of the United Nations. Retrieved from <http://www.mofa.go.jp/policy/economy/fishery/wsfs0911-2.pdf>
- Cotter, J., Davies, C., Nealon, J., & Roberts, R. (2010). Area Frame Design for Agricultural Surveys. *Agricultural Survey Methods*, 169–192. <https://doi.org/10.1002/9780470665480.ch11>

- de Bie, C. A. J. M. (2015). Remote Sensing and Geographic Information Systems Analysis Methods to Support Food and Water Security Studies (2014-2016). Enschede: University of Twente (Class Lecture Notes/Unpublished).
- Eurostat. (2015). Agricultural Census Introduction. Retrieved July 29, 2015, from http://ec.europa.eu/eurostat/statistics-explained/index.php/Agricultural_census_introduction
- FAO. (2010). Land Use Map of Rwanda. Huye: Food and Agriculture Organisation of the United Nations. Retrieved from <http://www.fao.org/3/a-au280e.pdf>
- FAO. (2016). *Enhancing Early Warning Capabilities and Capacities for Food Safety* (Training H). Rome: FAO.
- Funk, C., & Budde, M. E. (2009). Phenologically-tuned MODIS NDVI-based production anomaly estimates for Zimbabwe. *Remote Sensing of Environment*, 113(1), 115–125. <https://doi.org/10.1016/j.rse.2008.08.015>
- Holben, B. N. (1986). Characteristics of Maximum-value Composite Images from Temporal AVHRR Data. *International Journal of Remote Sensing*, 7(11), 1417–1434. <https://doi.org/10.1080/01431168608948945>
- Kathiresan, A. (2012). *Farm Land Use Consolidation in Rwanda: Assessment from the Perspectives of Agriculture Sector*. Kigali. Retrieved from http://www.minagri.gov.rw/fileadmin/user_upload/documents/agridocs/Farm_Land_Use_Consolidation_in_Rwanda.pdf
- Kennel, T. (2008). Area Frame. *Encyclopedia of Survey Research Methods*. Sage Publications, Inc. <https://doi.org/http://dx.doi.org/10.4135/9781412963947>
- King, M. D., Closs, J., Wharton, S., & Myers, M. (Eds.). (2003). *EOS Data Products Handbook (Volume 1)*. Greenbelt, Maryland: NASA Goddard Space Flight Center.
- Kiregyera, B., Megill, D., José, B., & Eding, D. (2007). Mission Report from a review of the National Agricultural Information System in Mozambique. Maputo: Instituto Nacional de Estatística. Retrieved from <https://www.dst.dk/ext/2170532/0/mozambique/MZ-2007-06-A-review-of-the-National-Agricultural-Information-System-in-Mozambique-by-Ben-Kiregyera-David-Megill-Bonifacio-Jose-and-Davi--pdf>
- Liew, S. C. (2001). Principles of Remote Sensing. Retrieved February 2, 2016, from <http://www.crisp.nus.edu.sg/~research/tutorial/rsmain.htm>
- LP DAAC. (2014). Vegetation Indices 16-Day L3 Global 250m. Retrieved June 23, 2015, from https://lpdaac.usgs.gov/dataset_discovery/modis/modis_products_table/mod13q1
- Luo, J., Ying, K., & Bai, J. (2005). Savitzky–Golay Smoothing and Differentiation Filter for Even Number Data. *Signal Processing*, 85(7), 1429–1434. <https://doi.org/10.1016/j.sigpro.2005.02.002>
- Maningas, R. V., Villagonzalo, S. O., & Macaraig, A. J. T. (2004). *Electronic Information Dissemination through the Farmers' Information and Technology Services (FITS)/Techno Pinoy Program: Bringing Information and Technology Within the Reach of the Farmers*. Laguna. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.202.7673&rep=rep1&type=pdf>
- Memarsadeghi, N., Mount, D. M., Netanyahu, N. S., & Le Moigne, J. (2007). A Fast Implementation of the ISODATA Clustering Algorithm. *International Journal of Computational Geometry & Applications*, 17(01), 71–103. <https://doi.org/10.1142/S0218195907002252>
- Menza, C., Caldow, C., Jeffrey, C., & Monaco, M. (2008). *Analysis of Sample Frames and Subsampling Methods for Reef Fish Surveys*. Retrieved from <http://ccma.nos.noaa.gov/publications/NCCOSTM72.pdf>
- MINIRENA, & RNRA. (2009). *Rwanda Aerial Photographs (25*25 cm)*. Kigali: Ministry of Natural Resources of Rwanda and Swedesurvey.
- NASA. (2002). *MODIS: Moderate-resolution Imaging Spectrometer*. Greenbelt, Maryland. Retrieved from http://modis.gsfc.nasa.gov/about/media/modis_brochure.pdf

- NASA. (2015). Measuring Vegetation (NDVI & EVI): Normalized Difference Vegetation Index (NDVI). Retrieved July 29, 2015, from http://earthobservatory.nasa.gov/Features/MeasuringVegetation/measuring_vegetation_2.php
- NISR. (2013). Seasonal Agricultural Survey Report: Rwanda. Kigali: National Institute of Statistics of Rwanda.
- NISR. (2015a). Seasonal Agricultural Survey (SAS) - 2015 Season A. Kigali: National Institute of Statistics of Rwanda.
- NISR. (2015b). Seasonal Agricultural Survey 2013 (Version 2). Kigali: National Institute of Statistics of Rwanda.
- NISR. (2015c). Seasonal Agricultural Survey 2014. Kigali: National Institute of Statistics of Rwanda.
- RNRA. (2010). Rwanda 2010 Classification Scheme 1. Kigali: Unpublished.
- Roser, M. (2015). Land Use in Agriculture. Retrieved July 29, 2015, from <http://ourworldindata.org/data/food-agriculture/land-use-in-agriculture/#note-3>
- Tucker, C. (2009). NOAA Advanced Very High Resolution Radiometer NDVI Data Set Comparison. Washington, D.C.: NASA/GSFC.
- Tucker, C. J., & Sellers, P. J. (1986). Satellite Remote Sensing of Primary Production. *International Journal of Remote Sensing*, 7(11), 1395–1416. <https://doi.org/10.1080/01431168608948944>
- Turner, M. G., & Gardner, R. H. (1994). *Quantitative Methods in Landscape Ecology: The Analysis and Interpretation of Landscape Heterogeneity*. Retrieved from https://books.google.nl/books/about/Quantitative_Methods_in_Landscape_Ecolog.html?id=8AuMildBNvIC&pgis=1
- Väisänen, P. (2009). *Integration of Registers and Survey-based Data in the Production of Agricultural and Forestry Economics Statistics*. Retrieved from <http://ec.europa.eu/eurostat/documents/1001617/4398397/S7P4-NEW-INTEGRATION-OF-REGISTERS-AND-SURVEY-BASED-DATA.pdf>
- Wigton, W. H., & Bormann, P. (1978). *A Guide to Area Sampling Frame Construction Utilizing Satellite Imagery*. New York.