Supporting Pro-Poor Reforms of Agricultural Systems in Eastern DRC (Africa) with Remotely Sensed Data: A Possible Contribution of Spatial Entropy to Interpret Land Management Practices

Pietro De Marinis 1,*, Samuele De Petris 2, Filippo Sarvia 2, Giacinto Manfron 1, Evelyn Joan Momo 2, Tommaso Orusa 2, Gianmarco Corvino 2, Guido Sali 3 and Enrico Mondino Borgogno 2

1 Department of Agricultural and Environmental Science, Production, Landscape and Agroenergy, University of Milan, 20133 Milano, Italy; giacinto.manfron@ec.europa.eu (G.M.); guido.sali@unimi.it (G.S.)
2 Department of Agriculture, Forest and Food Sciences, Torino University, 10124 Turin, Italy; samuele.depetris@unito.it (S.D.P.); filippo.sarvia@unito.it (F.S.); evelynjoan.momo@unito.it (E.J.M.); tommaso.orusa@unito.it (T.O.); g.corvino@genegis.net (G.C.); enrico.borgogno@unito.it (E.M.B.)

* Correspondence: pietro.demarinis@unimi.it; Tel.: +39-02-503-16466

Abstract: In the eastern Democratic Republic of Congo, agriculture represents the most important economic sector, and land control can be considered a perpetual source of conflict. Knowledge of the existing production system distribution is fundamental for both informing national land tenure reforms and guiding more effective agricultural development interventions. The present paper focuses on existing agricultural production systems in Katoyi collectivity, Masisi territory, where returning Internally and Externally Displaced People are resettling. We aim to define a repeatable methodology for building evidence-based and updated knowledge concerning the spatial distribution of the two existing production systems: subsistence-oriented agriculture (SOA) and business-oriented agriculture (BOA). To this aim, we used a supervised object-based classification approach on remotely sensed Sentinel-2 imagery to classify land cover. To classify production systems further within the “agriculture” and “pasture” land use classes, binary classification based on an entropy value threshold was performed. An iterative approach was adopted to define the final $H_{NDVI}$ threshold that minimised commission and omission errors and maximised overall accuracy and class separability. The methodology achieved acceptable observed accuracy (OA equal to 80–90% respectively for agricultural and pasture areas) in the assessment. SOA and BOA respectively covered 24.4 and 75.6% of the collectivity area (34,606 ha). The results conclude that land use and entropy analysis can draw an updated picture of existing land distribution among different production systems, supporting better-adapted intervention strategies in development cooperation and pro-poor agrarian land tenure reforms in conflict-ridden landscapes.

Keywords: eastern DRC; Masisi; agricultural production system; entropy analysis; Sentinel-2A; land use land cover

1. Introduction

The control of land has always been at the core of profound disputes in the Democratic Republic of Congo (DRC) [1–3]. The tragic series of events culminating in the Congolese war between 1994 and 2004 and the Tutsi Genocide, together with an ineffective state policy and a chronic insecurity situation, has led to the disruption of the agricultural sector and an increasingly intense conflict over access to land [4]. The historical background and varied orography of the Goma Diocese, located at the border between North and South Kivu provinces, has directly led to a complex agricultural landscape, with significant impacts on the natural environment and regional economic development [5,6]. The situation is further complicated by the massive number of Internally and Externally Displaced People...
(IDP and EDP, respectively); e.g., the OCHA estimated 863,400 IDPs between January 2009 and November 2014. According to existing international treaties, all these people have the right to return to their homes [7]. According to previous studies, socio-economic conditions profoundly influence terrestrial ecosystems, especially agroecosystems and their ability to support human settlements [8]. The agenda of international organizations dealing with peace-seeking and peacekeeping in the Eastern Congo now focuses on, among other issues, drawing a clear outline of land distribution among the different stakeholders in the agricultural sector [9,10]. Two different agricultural production systems exist in the region. These agricultural production systems are understood here as the two main types of livelihood strategies among the Congolese population of the Goma Diocese, namely Subsistence Oriented Agriculture (SOA) and Business Oriented Agriculture (BOA). The first is the type of agricultural activity defined as family farming by the FAO [11]. It is usually implemented on small plots of land, with access to very local markets. In DRC and specifically in the Goma Diocese, this production system forms a mosaicised land use/land cover (LULC) pattern that other authors identified as the “rural complex” [12,13]. The rural complex is a distinctive agricultural land cover mosaic surrounding the network of inhabited areas found along rivers and roads in DRC. It contains paths, grassy and bare communal areas, settlements and various land uses, primarily those associated with traditional smallholder livelihood shifting cultivation: cleared land, active fields, fallow fields, secondary forest and a permeable interface area with primary forest [12].

Conversely, huge plantations or grazing areas mostly characterise BOA, whose products are usually sold in foreign or regional markets. The two agricultural systems also differ in terms of land tenure, as SOA generally implies the exploitation of lands attributed to farmers through the traditional land tenure system and/or with no titling at all [11], while BOA usually relies on formally titled land. The two productive systems generate different costs and benefits for the local economy and existing socio-ecological systems [14–16]. In such a chronic emergency context, socio-economic development heavily depends on the resiliency of the agricultural sector [17]. Consequently, in-depth knowledge of regional agricultural geography is fundamental for understanding the spatial distribution and evolution of food supply chains, supporting the reintegration of IDPs/EDPs [18], and most of all assessing the impact of current and future policies and peacekeeping interventions [3,19,20].

To our knowledge, no official data exist regarding the spatial distribution of BOA and SOA systems in the Goma Diocese, especially at the scale of analysis of interest. In the DRC, and especially in the provinces of North and South Kivu, agriculture and livestock breeding constitute the backbone of socio-economic development [17]. Information regarding the distribution of land among the existing production systems is also fundamental for supporting advocacy actions towards the need for pro-poor land tenure reform in the DRC, which is debated since 2012 [21–25]. Given the characteristic inaccessibility of the territory, both in physical terms due to orography and insecurity, and in terms of data availability due to the lack of governmental official knowledge repositories, satellite remote sensing (RS) is a suitable option as an analysis tool. The present study aims to build evidence-based and updated knowledge regarding the spatial distribution of agricultural production systems in agricultural and pasture areas through land use analysis and entropy analysis of remote sensed Sentinel-2 imagery in the collectivity of Katoyi, North Kivu, DRC.

Although several previous studies tested the use of RS in the DRC [26–28], to the best of our knowledge, only two remote sensing-based LULC analyses focusing on the DRC exist [29,30], which present respectively a tiny scale of analysis and out-of-date information compared to our needs. Land use is strongly related to land cover, as the kind of activity implemented in a territory strongly depends on the features of the available environment and vice versa [31]. In the present study, we adopted the LULC acronym to focus on the functional definition of land use, which points to the description of the land in reference to its socio-economic purpose [32]. Therefore, LULC-change analyses can be of paramount importance for mapping the evolution of specific geographical contexts’ socio-economic
activities. This possibility is more significant where livelihood agricultural systems and forest preservation objectives represent coexisting aims.

LULC change studies are widely accepted to be essential for enabling preventive actions against natural resource degradation and destruction [33], especially in inaccessible zones such as the Congolese basin. As previously mentioned, LULC maps can highlight borders between different socio-economic uses within a given territory. Nevertheless, without other auxiliary information layers, it hardly describes the territory according to a series of operations carried out by humans in order to obtain specific products and/or benefits. This means that land use, in its sequential definition, cannot be inferred directly from land cover [34].

To unmix LULC mapping information within a specific LULC class, an entropy-based approach and texture analysis of satellite imagery seems to be promising and widely used in the literature [35–40]. Texture is an intrinsic property of virtual surfaces, and can be defined as the expression of patterns in the spatial variation of pixel values in imagery, which contain essential information concerning the structural arrangement of surfaces and their relationship to the surrounding environment [41]. There are different metrics used as expressions of textural features, and entropy is one of them. This parameter measures an image’s disorder: if an image is not uniform in terms of texture, a high entropy value will characterise it [42]. Entropy and other textural features have been widely used in remote sensing with reasonably satisfactory results [43]; in particular, these features can be directly used for scene classification or to improve LULC classification accuracy by adding information to spectrum-related data [44]. RS has significant potential for applications in this respect, but considering that this study area is located in the equatorial belt, some limitations should also be pointed out. On average, rainfall is very abundant in equatorial and forested regions characterised by intense evapotranspiration [45]. Moreover, high mountains characterise the study area and the surrounding region. For this reason, the area is known for the strong presence of clouds [46], which makes it challenging to collect multiple contiguous optical observations, both spatially and temporally [47–50].

2. Materials and Methods

2.1. Study Area

The present research focuses on the Katoyi Collectivity, located in Goma Diocese (GD), North Kivu, Democratic Republic of Congo (DRC). The GD (surface: 26,223 km$^2$) is located in the eastern DRC (DRC, Figure 1a,b). GD is an ecclesiastical administrative unit known as “small north” ("le petit Nord"). It is located in the North Kivu province and a small strip of the Kalehe Territory (Figure 1c) lying in the South Kivu province. Within the Diocese, there are 11 collectivities belonging to five territories (Goma, Rutshuru, Masisi, Walikale, and Kalehe). The study area is characterised by one of Africa’s highest population densities (2,211,000 inhabitants on 26,223 km$^2$ of territory, 88.4 inhabitants/km$^2$). A very high poverty level characterises the local population (72.9% of the population live below the UN poverty threshold in North Kivu, against 71.2% in the whole DRC) [51,52].
The GD’s orography is very complex. A continuous series of tiny plains, plateaus, hills and mountains make up the landscape. The local agricultural system can be framed in terms of the agroecological zones defined by FAO [53]: moving westward from the higher altitudes that lie along the borders with the neighbouring Uganda, Rwanda and Burundi and heading towards the forest, the farming system called “Highland Perennial” gives way to a “Forest-Based” system [53]. In this setting, the two previously mentioned production systems coexist. Small farmers mainly practice SOA, aiming at subsistence through family agriculture on small plots, mainly located near villages, on uplands officially under customary tenure or ad hoc agreements with private landowners. BOA is practised on or near large landholdings. It grows mainly cash crop plantations in monocultural cropping systems or breeds cattle in extensive systems [54]. Masisi territory and Katoyi collectivity, in particular, have been the theatre of massive population displacements from 1994 to 2004. During this period, the Katoyi collectivity hosted the headquarters of the leading armed group in the area at that time, the Democratic Forces for the Liberation of Rwanda [55]. Due to their presence, most of the population fled and left the collectivity scarcely inhabited during the following ten years. Since 2015, returning IDPs and EDPs have slowly resettled in the collectivity of Katoyi (according to unpublished CARITAS Development Goma reports of 2015).
2.2. Workflow

The present research firstly addresses the LULC of the GD in order to map the agricultural and pasture areas, and secondly assesses the existing agricultural production systems (i.e., SOA and BOA) in a strategic collectivity, i.e., Katoyi, in order to support rural development policies and programs. We review the proposed three-phase workflow in Figure 2.

![Figure 2. Workflow showing the main operations for obtaining final results in LULC acquisition from S2 multispectral imagery.](image)

During the preparation phase (see Figure 2), we collected and organised all relevant data in order to tackle spatial analysis by reducing elaboration time. During the object-based classification phase, we produced a LULC map of all of the GD starting from S2 multispectral images and Ground Reference Points (GRPs). In order to validate the classification, we used the Control Polygons (CPs) dataset.

During the third phase (see Figure 2), we addressed the classification of existing production systems in the Agriculture and Pasture classes. An NDVI entropy map was calculated for Katoyi collectivity, and classification of agriculture production systems was therefore produced through adopting an entropy threshold method at the patch level. To this aim, the CPs dataset was used to compute the confusion matrix and identify the threshold for the binary classification (SOA/BOA).

2.3. Preparation: Available Data

2.3.1. Sentinel-2 Data

Copernicus Sentinel-2 (S2) imagery was used to produce an up-to-date LULC of the area within this context. S2 is a wide-swath, high-resolution, multispectral imaging mission supporting Copernicus Land Monitoring studies, including vegetation, soil and inland waters. Spectral properties of the 13 bands acquired by the Multi-Spectral Instrument (MSI) are shown in Table 1. The nominal temporal resolution of S2 acquisition is five days; the geometric resolution (GSD, Ground Sampling Distance) is 10, 20 or 60 m depending on the band; the swath is 290 km². For this work, seven S2 Level-2A tiles were obtained from the Copernicus Open Access Hub geo-portal (scihub.copernicus.eu). Level-2A images
are supplied as ready-to-use, 100 km × 100 km tiles, ortho-projected in the WGS84 UTM reference frame [56] and calibrated at the Bottom of Atmosphere (BoA) reflectance level.

Table 1. Nominal features of S2.

<table>
<thead>
<tr>
<th>Name</th>
<th>Number of Bands</th>
<th>Sentinel-2A Central Wavelength (µm)</th>
<th>Sentinel-2B Bandwidth (µm)</th>
<th>Sentinel-2A + 2B Central Wavelength (µm)</th>
<th>Bandwidth (µm)</th>
<th>Spatial Resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blue</td>
<td>2</td>
<td>0.4966</td>
<td>0.098</td>
<td>0.4921</td>
<td>0.098</td>
<td>10</td>
</tr>
<tr>
<td>Green</td>
<td>3</td>
<td>0.56</td>
<td>0.045</td>
<td>0.559</td>
<td>0.046</td>
<td></td>
</tr>
<tr>
<td>Red</td>
<td>4</td>
<td>0.6645</td>
<td>0.038</td>
<td>0.665</td>
<td>0.039</td>
<td></td>
</tr>
<tr>
<td>NIR</td>
<td>8</td>
<td>0.8351</td>
<td>0.145</td>
<td>0.833</td>
<td>0.133</td>
<td></td>
</tr>
<tr>
<td>VegRedEdge</td>
<td>5</td>
<td>0.7039</td>
<td>0.019</td>
<td>0.7038</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>VegRedEdge</td>
<td>6</td>
<td>0.7402</td>
<td>0.018</td>
<td>0.7391</td>
<td>0.018</td>
<td></td>
</tr>
<tr>
<td>VegRedEdge</td>
<td>7</td>
<td>0.7825</td>
<td>0.028</td>
<td>0.7791</td>
<td>0.028</td>
<td></td>
</tr>
<tr>
<td>Narrow NIR</td>
<td>8a</td>
<td>0.8648</td>
<td>0.033</td>
<td>0.864</td>
<td>0.032</td>
<td></td>
</tr>
<tr>
<td>Swir1</td>
<td>11</td>
<td>1.6137</td>
<td>0.143</td>
<td>1.6104</td>
<td>0.141</td>
<td></td>
</tr>
<tr>
<td>Swir2</td>
<td>12</td>
<td>2.2024</td>
<td>0.242</td>
<td>2.1857</td>
<td>0.238</td>
<td></td>
</tr>
<tr>
<td>Coastal aerosol</td>
<td>1</td>
<td>0.4439</td>
<td>0.027</td>
<td>0.4423</td>
<td>0.045</td>
<td></td>
</tr>
<tr>
<td>Water Vapour</td>
<td>9</td>
<td>0.945</td>
<td>0.026</td>
<td>0.9432</td>
<td>0.027</td>
<td></td>
</tr>
<tr>
<td>Swir-Cirrus</td>
<td>10</td>
<td>1.3735</td>
<td>0.075</td>
<td>1.3769</td>
<td>0.076</td>
<td></td>
</tr>
</tbody>
</table>

Since the study area is located in the equatorial zone, cloud cover is not a negligible problem. Especially in forest and mountainous zones, intense evapotranspiration generates an almost constant cloud cover. Consequently, several scenes were obtained from the ESA SchiHub Geoportal [57], imposing a cloud cover threshold of 11% and ranging between March 2018 and November 2018. In order to limit residual cloud cover, each tile footprint was subsequently divided by selecting acquisitions in different periods resulting in 13 sub-tiles. Selected sub-tiles were then mosaicked using SAGA GIS v.7.0.0 [58]. Each subtitle is internally homogeneous from the spectral point of view, but different from adjacent ones that were possibly acquired on different dates. Nevertheless, the vegetation phenology at the equatorial zone is mainly the same throughout the year [59]. Names and locations of sub-tiles are shown in Table 2 and Figure 3.

Table 2. List of S2 sub-tiles used for this work.

<table>
<thead>
<tr>
<th>Subtile Name</th>
<th>Date (YYYY-MM-DD)</th>
<th>Cloud Cover %</th>
</tr>
</thead>
<tbody>
<tr>
<td>T35_MNT</td>
<td>2018_08_04</td>
<td>0</td>
</tr>
<tr>
<td>T35_MNU_N</td>
<td>2018_06_15</td>
<td>11</td>
</tr>
<tr>
<td>T35_MNU_S</td>
<td>2018_08_04</td>
<td>10</td>
</tr>
<tr>
<td>T35_MPT_E</td>
<td>2018_08_11</td>
<td>0.5</td>
</tr>
<tr>
<td>T35_MPT_W</td>
<td>2018_03_27</td>
<td>1.5</td>
</tr>
<tr>
<td>T35_MPU_NE</td>
<td>2018_11_19</td>
<td>1</td>
</tr>
<tr>
<td>T35_MPU_NW</td>
<td>2018_10_28</td>
<td>2.7</td>
</tr>
<tr>
<td>T35_MPU_SE</td>
<td>2018_11_19</td>
<td>1</td>
</tr>
<tr>
<td>T35_MPU_SW</td>
<td>2018_10_28</td>
<td>2.7</td>
</tr>
<tr>
<td>T35_MQT</td>
<td>2018_06_12</td>
<td>3</td>
</tr>
<tr>
<td>T35_MQU_N</td>
<td>2018_07_17</td>
<td>6</td>
</tr>
<tr>
<td>T35_MQU_S</td>
<td>2018_07_17</td>
<td>6</td>
</tr>
<tr>
<td>T35_MQV</td>
<td>2018_07_22</td>
<td>5</td>
</tr>
</tbody>
</table>
2.3.2. Open Datasets

Very High Resolution Satellite Imagery (VHRSI) was used for photo interpretation and validation/refinement of classification results, especially in those areas with clouds or shadows in any tile [60]. VHRSI provides real colour ortho-photos with a spatial resolution between 0.5 and 2 m, which are widely used in RS [61,62]. Specifically, web-accessible BING Aerial ortho-photos (updated 2018) were used [63]. Open Street Map (OSM) data were also used to map roads, water bodies and residential areas. We obtained OSM data in vector format from the online platform [64]. For the present research, we appropriately rasterised data at the 10 m grid size.

2.3.3. Reference Datasets

A ground survey campaign was performed throughout the entire diocese area to test the accuracy of LULC classification. One hundred GRPs were acquired using a Garmin GPSMAP 64 GNSS (Global Navigation Satellite System) receiver with a nominal positional accuracy of 5 m (Figure 4). For each GRP three data were acquired: planimetric position (according to the WGS84 UTM 35S GPS-acquiring Reference System–RS), a high-definition camera picture portraying the local LULC with annotated exposure, and the classification of local land use near that position into four main classes (i.e., Forest, Agriculture, Pasture and Non-vegetated areas). We then generated a reference polygon layer containing Control Polygons (CPs) to assess the classification of the agricultural production system in the Katoyi Collectivity. In particular, 200 CPs were selected randomly in Katoyi, based on the CP layer. One hundred CPs were selected from the “Agriculture” class as defined by the LULC classification; the remaining 100 CPs were selected from the “Pasture” class. Using VHRSI, a photo-interpretation of the CPs was performed via searching for SOA and BOA within the LULC classes. Figure 5 shows the CPs’ distribution in the study area.
2.4. Object-Based Classification in the Goma Diocese

An object-based classification (OBC) approach [65,66] was used for this work. This approach was based on the supervised classification carried out in four main steps: (a) S2 data segmentation; (b) training feature selection based on segmented objects; (c) object classification adopting a minimum distance algorithm; (d) cloud masking correction.

2.4.1. Segmentation

OBC divides a multi-spectral image into groups of spectrally homogeneous pixels while respecting some geometric constraints. Pixels from the processed image are consequently aggregated according to constraints (parameters) that the operator has to properly set up by considering the expected minimum object size and its maximum internal spectral heterogeneity. The segmentation step of the OBC approach was performed using the mean-shift algorithm available in Orfeo Toolbox v.6.6.0 open software [67,68]. The OBC algorithm aimed at minimising the spectral heterogeneity of polygons by comparing the
spectral properties of neighbour pixels. The resulting segmentation vector layer (SEG) was generated according to an a priori defined minimum mapping unit of 5000 m².

Segmentation was carried out with particular reference to the S2 bands having the finest GSD = 10 m, i.e., B2, B3, B4 and B8. Starting from the native BOA reflectance values, images were divided into segments with an internally homogeneous spectral response. Segments were then vectorised to generate the corresponding vector layer. During segmentation, required parameters were set to the values reported in Table 3.

### Table 3. Parameters used during image segmentation.

<table>
<thead>
<tr>
<th>Segmentation Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spatial radius</td>
<td>5 pixels</td>
</tr>
<tr>
<td>Range radius</td>
<td>150 DN</td>
</tr>
<tr>
<td>Mode convergence threshold</td>
<td>0.1 DN</td>
</tr>
<tr>
<td>Maximum number of iterations</td>
<td>100</td>
</tr>
<tr>
<td>Minimum region size</td>
<td>50 pixel</td>
</tr>
</tbody>
</table>

SEG was then used to explore internal features other than spectral signatures, such as recurrent radiometric patterns (texture) and shape. These were adopted in the second level of analysis to better interpret SEG meaning, and, for this work specifically, to investigate land fragmentation schemes possibly related to different agricultural management types.

#### 2.4.2. ROIs Selection

A false-colour image (FCI) was generated starting from mosaicked S2 bands, including B8, B4 and B3 in the channels of Red, Green and Blue respectively. Thirteen FCI classes were defined by searching for a specific land cover class, according to Appendix A, Table A1. Subsequently, 5491 regions of interest (ROIs) were randomly selected and photo-interpreted according to FCI classes. Figure 6 shows the ROIs’ locations. The total area of the ROIs corresponded to 1.6% of the entire surface of the GD. Table 4 reports the land cover classes adopted and the relative codes.

![Figure 6. Locations and number of pixels of ROIs (Reference Frame: WGS84/UTM35S).](image-url)
Table 4. Land cover codes.

<table>
<thead>
<tr>
<th>Land Cover Classes</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dense Forest</td>
<td>1</td>
</tr>
<tr>
<td>Sparse Forest</td>
<td>111</td>
</tr>
<tr>
<td>Open Forest</td>
<td>2</td>
</tr>
<tr>
<td>Bare Soil/Urban</td>
<td>3</td>
</tr>
<tr>
<td>Grassland and lower vegetation</td>
<td>4</td>
</tr>
<tr>
<td>Water</td>
<td>5</td>
</tr>
<tr>
<td>Clouds</td>
<td>6</td>
</tr>
<tr>
<td>Cloud Shadows</td>
<td>7</td>
</tr>
<tr>
<td>Lower vegetation with strong vigour</td>
<td>8</td>
</tr>
<tr>
<td>Agronomical field</td>
<td>9</td>
</tr>
<tr>
<td>Wetlands and burned areas</td>
<td>10</td>
</tr>
<tr>
<td>Fragmented croplands</td>
<td>11</td>
</tr>
<tr>
<td>Croplands</td>
<td>12</td>
</tr>
</tbody>
</table>

Figure 7 shows the average spectral signature of each ROI class according to the codes in Table 4. Since available S2 cloud masks could be ineffective in complete cloud detection, especially cirrus clouds [69], and considering the coarse geometric resolution of this mask (GSD = 60 m), to better map these classes, clouds and shadows were considered in the classification by adopting bands with a GSD of 10 m. In fact, Coluzzi [69] found that S2 cloud masks over rainforest areas had an omission error of more than 70%.

Figure 7. ROIs’ spectral signatures. Bold line represents the mean value; dotted lines are mean value ± 1 standard deviation. (a) Dense Forest-1; (b) Sparse Forest-111; (c) Open Forest-2; (d) Bare Soil/Urban-3; (e) Grassland and lower vegetation-4; (f) Water-5; (g) Clouds-6; (h) Cloud Shadows-7; (i) Lower vegetation with strong vigour-8; (j) Agronomical field-9; (k) Wetlands and burned areas-10; (l) Fragmented croplands-11; (m) Croplands-12.
2.4.3. “Minimum Distance” Classification

A supervised OBC relies on a minimum distance algorithm approach. Different works have highlighted the performance and capability of such a classifier when applied to agricultural case studies [70–72].

Exploiting the B2, B3, B4 and B8 S2 bands, the relative mean value was computed at the patch level and added to the attribute table, starting from the SEG layer. Each patch’s spectral signature was classified based on ROI training features, adopting the minimum distance algorithm tool available on SAGA GIS v. 7.0.0. The land cover map was finally created, adopting the previously mentioned class codification.

2.4.4. Cloud Masking Correction

As the land cover classification was still affected by cloud cover contamination, the patches classified as clouds and cloud shadows were selected and substituted with the real land cover class retrieved from VHRSI photo-interpretation.

2.4.5. Land Use Class Aggregation

Thirteen land cover classes characterised the map obtained by the classification. These classes were initially chosen to better cope with the spectral variability among all the land cover types featured in the study area, but to restrict the focus on vegetation-related LULC classes, fewer classes were sufficient. Therefore, classes were merged as indicated in Table 5 to finally obtain a five land-use classification consisting of forests (coded as 1), agricultural (coded as 2), pasture (coded as 3), non-vegetated (coded as 4), and water (coded as 5). The class “water” was excluded in the successive validation step, as it was not relevant to the aim of this work.

Table 5. Land cover and LULC codes adopted.

<table>
<thead>
<tr>
<th>Land Cover Codes</th>
<th>LULC Codes</th>
<th>LULC Class Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1; 2; 111</td>
<td>1</td>
<td>Forests</td>
</tr>
<tr>
<td>9; 11; 12</td>
<td>2</td>
<td>Agricultural</td>
</tr>
<tr>
<td>4; 8</td>
<td>3</td>
<td>Pasture</td>
</tr>
<tr>
<td>3; 10</td>
<td>4</td>
<td>Non-vegetated</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>Water</td>
</tr>
</tbody>
</table>

The accuracy of the LULC map was tested against the GRPs by comparing classes defined by ground-based methods with those classified by the minimum distance algorithm. A confusion matrix was then produced to evaluate both classification performances.

2.5. Classification of the Production Systems in Katoyi

In this work, NDVI entropy ($H_{NDVI}$) was assumed as an appropriate image texture parameter able to separate the agricultural production systems of interest. In fact, building on the description of the two existing production systems, we assumed that in BOA production systems the vegetated areas (patches) would show a more homogeneous NDVI distribution, corresponding to lower $H_{NDVI}$ values. Conversely, in SOA production systems vegetated areas would be more heterogeneous and fragmented, making $H_{NDVI}$ of these patches higher. With these premises, the most critical aspect in the workflow was the identification of an appropriate $H_{NDVI}$ threshold able to separate the two production systems (i.e., SOA and BOA). Obviously, an objective and generally viable threshold does not exist. Consequently, an adaptive method was implemented, which was based on an iterative process, looking for the $H_{NDVI}$ value that minimises classification commission and omission errors while maximising the corresponding overall accuracy and class separability.
To this aim, an NDVI map [73] was calculated to describe vegetation biomass starting from multispectral mosaics, shown in Equation (1):

$$NDVI = \frac{\rho_{NIR} - \rho_{RED}}{\rho_{NIR} + \rho_{RED}}$$ (1)

where $\rho_{NIR}$ and $\rho_{RED}$ are S2 B8 and B4 reflectance values, respectively.

$H_{NDVI}$ index was then calculated according to Equation (2) to assess local vegetation heterogeneity [41]:

$$H_{NDVI} = -\sum_{i=1}^{N} \sum_{j=1}^{N} NDVI_{i,j} \log(NDVI_{i,j})$$ (2)

where $NDVI_{i,j}$ is the $NDVI$ value at the $i$th row and $j$th column in the local square window with a size of $N$ pixels. For this study, the authors adopted a window size of $10 \times 10$ pixels. Finally, an $H_{NDVI}$ map was generated. The average $H_{NDVI}$ value was calculated at the patch level of the SEG layer. Only patches belonging to agriculture and pasture classes were considered in order to better characterise these LULC types. To do this we adopted the following interpretative key: Low $H_{NDVI}$ values meant homogeneous vegetated surfaces; high $H_{NDVI}$ meant highly heterogeneous vegetation, probably mixed with bare soil, or coexistence of vegetation covers with variable height and density [74].

Two binary classifications based on an entropy value threshold were performed for agriculture and pasture classes, respectively. The confusion matrix parameters were calculated according to Richards et al. [75] and using the CPs as references. In particular, class separability was estimated using the Jeffries–Matusita index ($JM$) according to Equation (3):

$$JM = 2\left(1 - e^{-B}\right),$$ (3)

in which

$$B = \frac{1}{8} (m_i - m_j)^T \left(\frac{C_i + C_j}{2}\right)^{-1} (m_i - m_j) + \frac{1}{2} \ln \left(\frac{|C_i+C_j|}{|C_i||C_j|^{1/2}}\right)$$ (4)

where $C_i$ and $C_j$ are the within-class covariance of the $i$th and $j$th classes respectively, and $m_i$ and $m_j$ are the means of the $i$th and $j$th classes respectively. A JM distance equal to 2.0 implies 100% separability between two classes [76]. Therefore, 10 classifications were generated for agriculture LULC based on the following $H_{NDVI}$ threshold values: 0.3; 0.4; 0.5; 0.6; 0.7; 0.8; 0.9; 1.0; 1.1; 1.2; while eight classifications were generated for pasture LULC based on thresholds of 0.3; 0.4; 0.5; 0.6; 0.7; 0.8; 0.9; 1.0. The binary classification minimising commission and omission errors and maximising overall accuracy and JM was assumed to be the most reliable for classifying agriculture and pasture production systems.

3. Results and Discussion

3.1. LULC of the Goma Diocese

The entirety of the GD was divided into four classes of LULC: forest, agricultural, pasture and non-vegetated areas (Figure 8). Roads and water bodies were obtained from OSM data and overlapped on the LULC map, since they were not relevant for the work focused on vegetation-related classes.

As shown in Table 6, the most widespread class was forest, with an extent of 15,629.13 km$^2$ corresponding to 59% of the GD. The agricultural class covered 5106.25 km$^2$ and the pasture class 3593.71 km$^2$, corresponding to shares of 19% and 13%, respectively. By observing the maps, it is possible to note that the areas containing forests were mainly located in the central-western part of the GD, while agricultural and pasture land were more represented in the central-eastern zone.
Table 7 shows the confusion matrix related to the LULC classification produced in the GD. The classification performance resulted in an overall accuracy (OA) of 0.55; producer accuracy (PA) and user accuracy (UA) values (Table 7) are helpful to better clarify the strengths and weaknesses of this result.

In particular, the classification was accurate in recognising “forest and urban” and “bare soil” classes, since their spectral signatures were clearly separable from the others. On the other hand, there was an overlap between agricultural and pasture classes, since these have a similar spectral signature, which makes their separation difficult. A multi-temporal approach can investigate vegetation’s phenological behaviour, improving the separation between these classes [77]. Unfortunately, the adopted approach (based on single mosaicked acquisition) was dictated by the intense cloudiness that generally characterises equatorial areas, making it impossible to collect a series of contiguous images in space-time.

Table 6. Extent of LULC classes in GD.

<table>
<thead>
<tr>
<th>LULC Classes</th>
<th>Area (km²)</th>
<th>Area (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forest</td>
<td>15,629.13</td>
<td>59.02</td>
</tr>
<tr>
<td>Agricultural</td>
<td>5106.25</td>
<td>19.28</td>
</tr>
<tr>
<td>Pasture</td>
<td>3593.71</td>
<td>13.57</td>
</tr>
<tr>
<td>Non-vegetated areas</td>
<td>1997.12</td>
<td>7.54</td>
</tr>
<tr>
<td>Water</td>
<td>153.39</td>
<td>0.59</td>
</tr>
<tr>
<td>Total</td>
<td>26,479.62</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 7 shows the confusion matrix related to the LULC classification produced in the GD. The classification performance resulted in an overall accuracy (OA) of 0.55; producer accuracy (PA) and user accuracy (UA) values (Table 7) are helpful to better clarify the strengths and weaknesses of this result.

Table 7. Confusion matrix of LULC. UA = User Class Accuracy, PA = Producer Class Accuracy, CE = Commission Error, OE = Omission Error.

<table>
<thead>
<tr>
<th>Class from LULC (n° Pixels)</th>
<th>Forest</th>
<th>Agriculture</th>
<th>Non-vegetated areas</th>
<th>Pasture</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forest</td>
<td>24</td>
<td>4</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Agricultural</td>
<td>17</td>
<td>17</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Non-vegetated areas</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td>Pasture</td>
<td>9</td>
<td>4</td>
<td>0</td>
<td>6</td>
</tr>
</tbody>
</table>

Metrics (%)

<table>
<thead>
<tr>
<th>LULC classes</th>
<th>UA</th>
<th>PA</th>
<th>CE</th>
<th>OE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forest</td>
<td>48</td>
<td>82.8</td>
<td>17.2</td>
<td>52</td>
</tr>
<tr>
<td>Agriculture</td>
<td>68</td>
<td>41.5</td>
<td>58.5</td>
<td>32</td>
</tr>
<tr>
<td>Non-vegetated areas</td>
<td>81.8</td>
<td>81.8</td>
<td>18.2</td>
<td>18.2</td>
</tr>
<tr>
<td>Pasture</td>
<td>42.9</td>
<td>31.6</td>
<td>68.4</td>
<td>57.1</td>
</tr>
</tbody>
</table>
In particular, the classification was accurate in recognising “forest and urban” and “bare soil” classes, since their spectral signatures were clearly separable from the others. On the other hand, there was an overlap between agricultural and pasture classes, since these have a similar spectral signature, which makes their separation difficult.

A multi-temporal approach can investigate vegetation’s phenological behaviour, improving the separation between these classes [77]. Unfortunately, the adopted approach (based on single mosaicked acquisition) was dictated by the intense cloudiness that generally characterises equatorial areas, making it impossible to collect a series of contiguous images in space-time.

3.2. Production System Identification in Katoyi Collectivity

The binary classifications for agriculture and pasture production systems were defined according to the confusion matrix and JM parameters (Figure 9), searching for the $H_{NDVI}$ threshold value that minimised Omission Error (OE) and Commission Error (CE) for both classes and at the same time maximised the OA and the JM values. The $H_{NDVI}$ value of 0.7 was identified as the best for both the LULC classes (Figure 9).

![Figure 9. Accuracy parameters used to select the NDVI entropy threshold ($H_{NDVI}$, red vertical line) to generate a binary classification of rural production systems. (a) Agriculture LULC accuracy. (b) Pasture LULC accuracy. Omission and Commission Errors (OE and CE) produced during Business Oriented Agriculture systems classification (OE-BOA and CE-BOA) are marked by solid and dotted orange lines, respectively. Omission and Commission Errors generated during Subsistence Oriented Agriculture systems classification (OE-SOA and CE-SOA) are marked by solid and dotted green lines. The overall accuracy (OA) is represented by solid black lines and the Jeffries–Matusita (JM) values are depicted by blue lines.](image)

By adopting this threshold, the OA and JM of agriculture LULC were equal to 90% and 1.24, while OE and CE were lower than 15% for both BOA and SOA. Concerning pasture LULC, OA and JM were equal to 80% and 1.22 while OE and CE were lower than 30% for both BOA and lower than 20% for SOA. All other tested thresholds in the range 0.3 < $H_{NDVI}$ < 1.2 showed lower OA and higher OE/CE.

It is worth remembering that an interpretation of $H_{NDVI}$ behaviour is needed to determine the meaning of classes. This interpretation was achieved by looking for $H_{NDVI}$ class distributions (Figure 10). High entropy in the agriculture class denoted strong spectral variability within each patch, probably related to bare soil/vegetation alternation typical of SOA; inversely, in the BOA a low variability index of vegetation was expected due to homogeneous management of the surface. Conversely, for the pasture class the meaning of entropy changed; in fact, in a pasture system the rate use of vegetation follows grazing pressure [78,79]. High grazing pressure generates a homogeneous consumption of pasture, thus a low $H_{NDVI}$ value within each patch. SOA is characterised by small grazing surfaces and intensive local use of pastoral resources. Ordinarily, a BOA pasture system has
large surfaces (low grazing pressure) [80], which results in heterogeneous consumption of pasture and increasing \( H_{NDVI} \) values.

![Boxplots of \( H_{NDVI} \) for agriculture and pasture LULC classes](https://example.com/figure10.png)

**Figure 10.** \( H_{NDVI} \) boxplots for Katoyi agriculture and pasture LULC classes. (Orange) BOA systems; (Green) SOA systems.

With this interpretative key, a final production systems map (Figure 11) was generated as follows: in the agriculture class, all patches with average \( H_{NDVI} \geq 0.7 \) were classified as SOA, while patches with \( H_{NDVI} < 0.7 \) were classified as BOA. In the pasture class, the same principle was applied in inverse.

![Map of agriculture and pasture production systems in Katoyi](https://example.com/figure11.png)

**Figure 11.** Map representing agriculture and pasture production systems in Katoyi (Reference Frame: WGS84/UTM33S).

The total hectares ascribed to BOA and SOA in Katoyi collectivity are reported in Table 8. These results proved the prevalence of BOA production systems compared to SOA in Katoyi, and seem to confirm the portrait of a region where land scarcity for smallholders is becoming more and more relevant, especially where the dispute between large landowners and subsistence farmers is a perpetuating factor of conflict [2]. This result seems to agree with the unpublished results of a survey, conducted in 2017 by
Caritas Development Goma, estimating that large landowners control approximately 80% of the land in the GD. On the other hand, the validity of the methodology and the obtained accuracy in classifying the existing production systems could not be compared with those of other studies, due to the lack of research applying the same methodology in this region [81]. To the best of our knowledge, the only similar studies focus on different regions and address classification through different methodology, built on different available data, while obtaining comparable accuracy and methodological validity [82,83].

Table 8. Agriculture and pasture production system areas in Katoyi according to the LU map produced.

<table>
<thead>
<tr>
<th>LULC Classes</th>
<th>Production System</th>
<th>ha</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agriculture</td>
<td>BOA</td>
<td>80%</td>
</tr>
<tr>
<td></td>
<td>SOA</td>
<td>20%</td>
</tr>
<tr>
<td>Pasture</td>
<td>BOA</td>
<td>68%</td>
</tr>
<tr>
<td></td>
<td>SOA</td>
<td>32%</td>
</tr>
</tbody>
</table>

The proposed methodology and the specific approach for determination of an appropriate entropy threshold value were applied to increase the procedure’s general applicability. In fact, both the adoption of publicly available data and the introduction of processing steps specifically aimed at reducing/avoiding subjective decisions were the basis for our work. This made it possible to calibrate $H_{NDVI}$ threshold value depending on the local distribution and value of $NDVI$ at the patch scale, both of which strictly depend on local land management. Consequently, the $H_{NDVI}$ threshold values reported in this work cannot be assumed a priori but need to be determined empirically. However, the approach for their determination is general and adaptable and, therefore, usable anywhere around the world provided that ground reference data (here CPs) are collected to make the computation of classification errors possible.

4. Conclusions

In this work, S2 multispectral images were used to assess LULC in the GD. Specifically, a first land cover map was produced using an object-based supervised classification approach to detect and characterise land use in the GD. Comparable classification accuracy was found in a previous study [84], reinforcing S2’s data reliability when applied in the African context. Furthermore, in line with past research experience [66,85], the joint use of S2 and segmentation procedures proved to be effective in mapping and characterising agricultural and forest contexts. Nevertheless, some problems persisted. For instance, in SOA, where small heterogeneous areas were present and the algorithm could not systematically recognise fields, the delineation proved to be highly sensitive to segmentation parameters.

Starting from the identified LULC agriculture and pasture classes, a classification of agriculture production systems was created based on $H_{NDVI}$ value at the patch level. Since we presumed only two existing production systems (i.e., BOA and SOA), we proposed a binary threshold classification method. Such an approach is undoubtedly affected by threshold selection, and many works have examined this problem [86,87]. In this work, we proposed a retrospective method to define the best threshold and generate a binary classification. This threshold was defined by an optimisation procedure based on confusion matrix parameters and class separability. Specifically, we found that 0.7 $H_{NDVI}$ was the threshold that minimised CE and OE and maximised OA and class separability (the JM index) simultaneously for both agriculture and pasture. Some critical points still persist here: (a) more reference data (in this work, partially produced by photo interpretation) are needed to calculate confusion matrix parameters; (b) imbalanced data in the binary classification could affect results [88,89]; (c) further interpretation is needed to correct class meaning in order to avoid the introduction of bias. Nevertheless, we found that correcting class meaning through the $H_{NDVI}$ distribution between classes was a useful tool for classifying agriculture production systems; OA was found to equal 90% and 80% for agriculture and pasture, respectively.
Overall, the obtained results led us to conclude that RS was a useful tool for spatial analysis of land distribution among different agricultural production systems, even in equatorial regions characterised by complex orography and wide agricultural landscapes. This is extremely interesting in a region such as North Kivu and particularly the GD, where digital cadastres do not exist and the orography and chronic security issues result in difficult access to land-related information. In the framework of the Congolese path towards pro-poor agrarian land tenure reform that should regulate the increasing trend towards land ownership concentration, and in the context of an increasing rural population that strives for access to land, this information is fundamental for supporting both evidence-based interventions at national policy level, and international development cooperation operating within the country.

Research Perspectives

In this study, we obtained interesting results on a pilot-area scale (Katoyi collectivity); therefore, it would be very interesting to scale up this approach to the whole GD area. Across the world, there are several examples of countries where pro-poor land tenure reforms are struggling to see light due to specific, often-inscrutable political, security and geographical contexts. Knowledge regarding the distribution of different production systems may be the key information to trigger or support pro-poor reforms. Several cases exist in Africa where the proposed methodology may be trialled and therefore further improved and validated. Specifically, it is worth highlighting that a more robust validation of the proposed method could be obtained by repeating the implementation in several other case studies and by using ground-reference data systematically. The repetition in such contexts would fall under a Science Diplomacy approach [90], which uses applied research to foster diplomacy and social innovation uptake concerning delicate issues such as rural reforms in the framework of the 2030 Agenda.

Among the main critical issues to be improved, the region-specific constraint regarding cloud coverage could be overcome by coupling multispectral passive remote sensing with active remote sensing SAR (Synthetic Aperture RADAR) data (e.g., Sentinel-1). In fact, SAR sensors can penetrate cloud cover, and the polarisation properties of the backscatter signal and its temporal behaviour could more extensively describe land use classes. SAR data would also facilitate the adoption of a multi-temporal approach to the analysis of LULC and production systems, possibly improving the classification accuracy.

Nowadays, new software and algorithms such as Google Earth Engine (GEE) can improve remotely sensed data processing and allow data computation for large areas.

Future developments may concern the joint use in GEE of SAR and optical data, and the exploitation of alternative classification approaches (e.g., Random forest or TensorFlow).

Finally, a pixel-based approach to the analysis of VHRSI could also be tested and compared to the proposed methodology to assess the margin of improvement in classification accuracy.


Funding: This research was funded by the ARDST project within the EU strategy for peace seeking and keeping operations in the African Great Lakes region (Grant No. 360149 of ICSP–Instrument contributing stability and peace).

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to credit restrictions.

Acknowledgments: The framework for this study is provided by a three-year EU-funded project called ARDST “Appui au retour de réfugiés et déplacés par le biais de la sécurisation de terres en
Diocèse de Goma”, led by Caritas Development Goma NGO. The project started in February 2016 and ended in June 2019. We would like to acknowledge the fundamental role of the whole ARDST project team, particularly Gabriel Habimana, Janvier Kashori, Henri Kambesa, Claude Sibomana Ntwari, Joseph Habaryrimana Niangi, Robert Akilimali, Joseph Mvuyekure, Théoneste Buzakare, Guido Semahane, Serge Kwisila, Milabyo Malekela, Bertin Kibanja Mishiki, Christophe Sibomana, Michel Iyanya and Patrick Nsengiyumva.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Appendix A. Coding Adopted for Photointerpretation of FCI

Table A1. False-colour image (FCI) from mosaicked S2 bands, used to create thirteen land cover classes used in turns to define Regions of Interest.

<table>
<thead>
<tr>
<th>Land Cover Classes</th>
<th>Codes</th>
<th>FCI Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dense Forest High Vigour</td>
<td>1</td>
<td><img src="image1" alt="Image" /></td>
</tr>
<tr>
<td>Dense Forest Low Vigour</td>
<td>111</td>
<td><img src="image2" alt="Image" /></td>
</tr>
<tr>
<td>Open Forest</td>
<td>2</td>
<td><img src="image3" alt="Image" /></td>
</tr>
<tr>
<td>Bare soil/urban</td>
<td>3</td>
<td><img src="image4" alt="Image" /></td>
</tr>
<tr>
<td>Sparse vegetation low vigour</td>
<td>4</td>
<td><img src="image5" alt="Image" /></td>
</tr>
</tbody>
</table>
Land Cover Classes | Codes | FCI Example
--- | --- | ---
Low vegetation—high vigour | 8 | ![Image](image1.png)
Ploughed fields | 9 | ![Image](image2.png)
Wet or burned areas | 10 | ![Image](image3.png)
Fragmented agricultural areas | 11–12 | ![Image](image4.png)

**Note:** Within the framework of the ARDST Project, we supervised the implementation of a survey targeting the holders of land concessions in the Goma Diocese, aiming to collect data regarding existing conflicts concerning access to land. The survey allowed for the collection of 2558 observations. Surveyed concessions covered 173,526.3 ha (equal to 1735.26 km², on 8,609.96 km² of agricultural and pasture area in GD). Concessions greater than fifty hectares comprised a share of 84% of the surveyed area.

**References**


52. World Poverty Clock.
60. OSM. Open Street Map. 2020. Available online: https://www.openstreetmap.org/#map=4/38.01/-95.84 (accessed on 8 December 2021).
70. OSM. Open Street Map. 2020. Available online: https://www.openstreetmap.org/#map=4/38.01/-95.84 (accessed on 8 December 2021).
70. Wacker, A.G.; Landgrebe, D.A. *Minimum Distance Classification in Remote Sensing*; LARS Technical Reports; Purdue University: Lafayette, IN, USA, 1972.


85. Immitzer, M.; Vuolo, F.; Atzberger, C. First Experience with Sentinel-2 Data for Crop and Tree Species Classifications in Central Europe. *Remote Sens.* 2016, 8, 166. [CrossRef]


