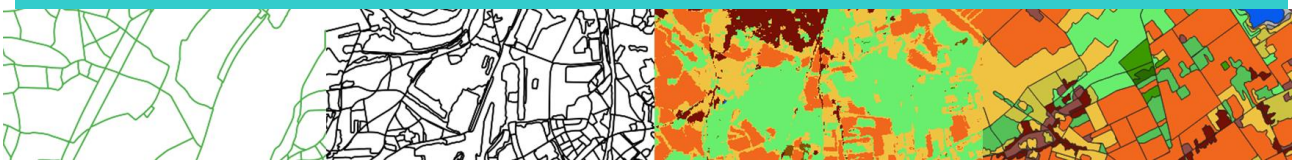
Service contract for the Copernicus Land monitoring services

Crop Mapping for GEOGLAM Country Level Support



Framework contract 939708-2020-IPR

End-of-season Crop Type Map & Crop Mask
Kenya - long rains season – 2023

Prepared by:

CLS
COLLECTE LOCALISATION SATELLITES

&

TerraSphere

Reference : End-of-season mapping - Kenya - long rains season 2023

Issue 2.1 - 29/11/2023

Limited distribution/Diffusion limitée

Parc de la Cimaise I, 27 rue du Carrousel, 59650 Villeneuve d'Ascq, France

Tel +33 (0)3 20 72 53 64 Fax +33(0)3 20 98 05 78

www.sirs-fr.com

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Issues :

1.0	09-11-2023	First version submitted to JRC
2.0	22-11-2023	Second version submitted to JRC
2.1	29-11-2023	Minor corrections

1 Introduction

This document describes the end-of-season mapping of the crop type and crop mask for an Area Of Interest (AOI) in Kenya during the long rains season in 2023. The AOI is extended from 99,000 km² (applied in 2021) to 181,000 km² for the current season. This document summarizes the workflow and any methodological change (put in place to obtain the above-mentioned products) with respect to what was described in the feasibility study and conducted during the previous long-, and short rains seasons. The document also describes the satellite imagery, the fieldwork and the ground truth data used for the classification, satellite data pre-, and post-processing to get an understanding of the workflow used. Where changes have been made from previous seasons, these are described in more detail in this report, together with the reasons for the changes.

Unlike previous rainy seasons, Kenya's current long rains season hasn't been severely affected by drought[1] as shown in Figure 1**.** Abundant rainfall has been recorded in the AOI since the beginning of the season, which has been beneficial for the long rains. Vegetation conditions in most of the country have improved compared to previous months and are above average, except for parts of the coastal areas and Taita Taveta county which show early negative anomalies and had experienced a late start of the season. So, in the south-east of the AOI, initial crop growth was delayed by the drought. Reports indicate that the long rains season crops in the main producing areas of western and central Kenya generally performed well in the latter part of the season, thanks to sustained rainfall from the early part of the season. Consequently, no major drought conditions warning was visible in the country.

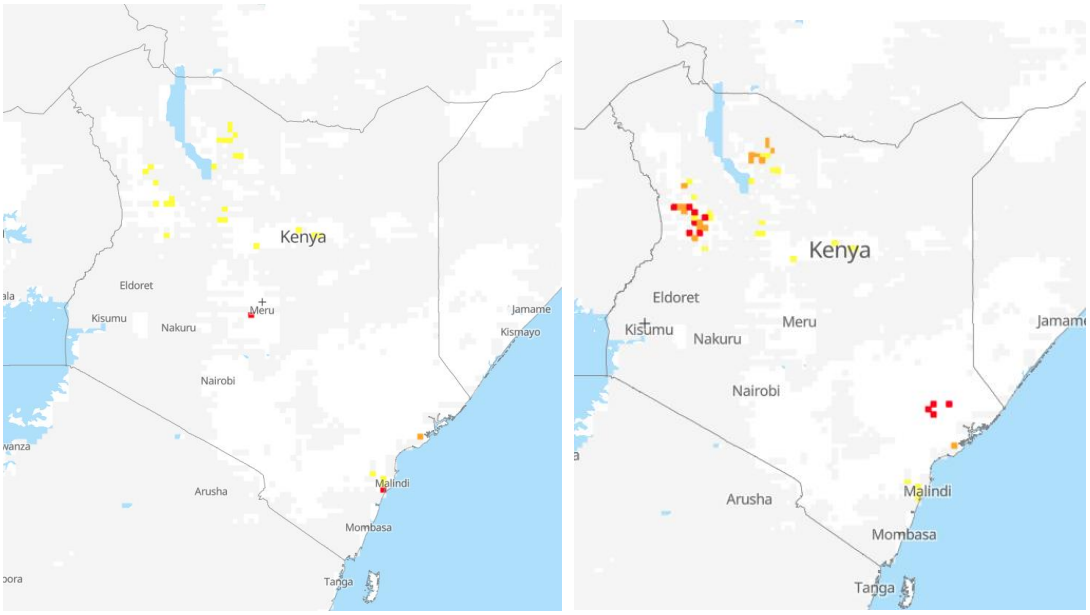


Figure 1: Seasonal Combined Drought Indicator – April-June (left) June-August 2023 (right) (Source: East Africa Drought Watch https://droughtwatch.icpac.net/mapviewer/)

[1] https://agricultural-production-hotspots.ec.europa.eu/country.php?cntry=133

2 Summary of methodology

This section provides a summary of the methodology used by the consortium to derive the sample units used in the implemented workflow as training and validation data to produce the crop type and the crop mask maps. The methodology is also described in the feasibility study (D1.1) and Annex I, Annex II and Annex III provide background information on the stratification and the sampling design.

In order to account for to variability and have a statistically valid product, a stratified sampling approach is proposed for the implementation of the field campaign and the provision of training and validation data. This approach consists of the following steps:

1. **Definition of a suitable stratification:** identify homogeneous strata with respect to agricultural practices and cropping systems and identify areas without agriculture activities.
2. **Selection of an appropriate sample unit**: define an appropriate type of sample unit to optimise the field data collection while guaranteeing the acquisition of sufficient high-quality data.
3. **Adoption of a suitable sampling strategy**: the purpose is to guarantee that the data collected accurately represents the various conditions in the selected AOI.
4. **Identification of the sample size per stratum**: this involves drawing the final sample content and sample size per stratum.

In order to meet the requirements of the applicants the following details needed to be captured; 1) all land cover types including the pre-defined main crop type of applicants interest, 2) separating irrigation from rainfed fields and 3) presence of mixed cropping.

2.1 Stratification

After analysing the AOI, a stratification with 6 strata is chosen. These strata were defined based on 3 different discriminating factors: 1) the expected land cover, 2) the agro-ecological zone and 3) the elevation.

Even though the data is older (dating from 1995-2002), the source chosen for the landcover is the FAO's AFRICOVER dataset. It provides greater spatial details in comparison to other dataset, and enables the distinction between rainfed and irrigated cropland. The WRI dataset will allow to differentiate the two crop systems in separate strata and rice fields**,** cultivated through irrigation most of it, will be included in the irrigation crop strata.

For the second factor, the Agro-Ecological Zones (AEZ) for Africa South of the Sahara[2] dataset has been chosen. It is a dataset with 10 km resolution for the reference year 2015 developed based on the methodology developed by the FAO and the International Institute for Applied Systems Analysis (IIASA). For Kenya, 4 main AEZ are determined with one class having two sub-types resulting in 5 classes; 1) Humid, 2) Sub-Humid, 3) Semi-Arid, 4a) Tropical Highlands Humid and 4b) Tropical Highlands Sub-Humid.

Finally, the Digital Elevation Model[3] has been used, splitting the elevation into 3 classes: 1) 0-1000 meters as lowlands, 2) 1000-3000 meters as highlands and 3) all above 3000 m as not suitable for agriculture.

[2] https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/M7XIUB

[3] https://spacedata.copernicus.eu/web/cscda/dataset-details?articleId=394198

The 3 above factors have been combined to constitute the following 6 strata; 1) Irrigated crops (including rice fields), 2) Rainfed Lowlands, 3) Rainfed Highlands Sub-Humid/Humid, 4) Rainfed Tropical Highlands Humid, 5) Rainfed Tropical Highlands Sub-Humid and 6) Other areas (See Figure 2).

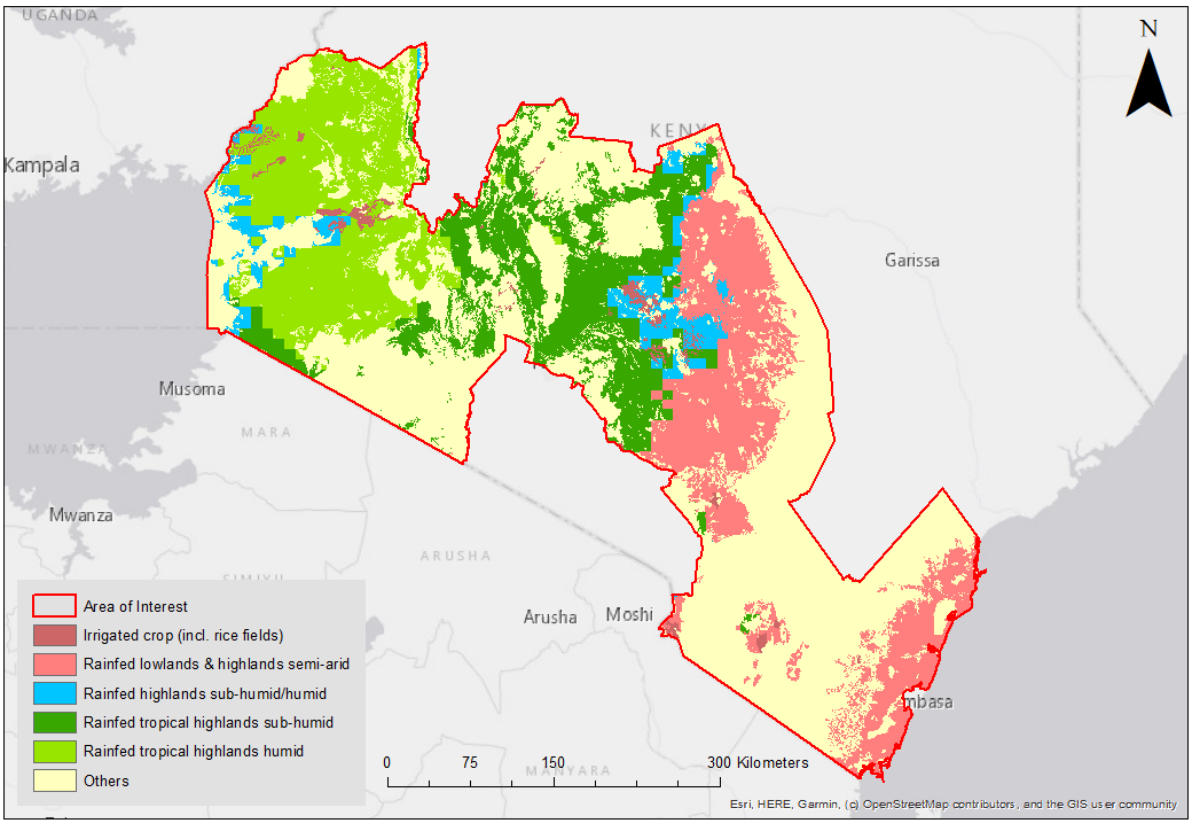


Figure 2 Derived AOI stratification

2.2 Sample design

To sample the whole AOI, a square segment design is implanted. Compared to point sampling, the use of segments has several advantages: 1) the entire area is described and 2) land cover characteristics can be extracted based on their proportion.

A two-stage sampling approach is implemented by 1) applying a 20 x 20 km grid over the AOI and 2) subsequently selecting a sample unit within each grid cell. The size of the square segments should be determined based on two criteria: 1) optimising the level of effort required to collect the data in the field and 2) maintain the required level of precision. Based on previous experience documented in the literature (Taylor, 1997), it is recommended that a segment should not contain more than 10 to 30 field parcels on average. This is to ensure a good compromise between the total number of segments, the time required to survey each segment and the spatial distribution of these sample units. Based on county information provided by the JRC, the crop field parcels typically range from 1 to 5 acres. Considering an average of 3 acres (*i.e.*, 1.2 ha), this would correspond to a total of 20 crop field parcels (assuming that the entire segment is covered with crops) a segment size should cover 25 ha. To have sensible dimensions and considering that most segments also contain other land use we suggest a segment with a unique size of 500 x 500 m and

covering 25 ha. This assumption was tested and confirmed to be correct by digitizing these segments and counting the cropland parcels within.

The final selection of segments is done via **stratified systematic random sampling** in order 1) to gather training data for the identification of land cover and crop types and 2) to gather data for the provision of unbiased crop area estimates and the validation of the crop type maps and crop mask.

The first stage of the approach is implemented by applying a grid over the overall area of the AOI and ensures that the segments that are selected are distributed over the entire AOI. The second stage refers to the selection of sample units within each grid cell. The grid cell is often referred to as a block (Gallego, 1995). Two-stage sampling is considered suitable for accuracy assessment of land cover maps and area estimation of land cover types (Stehman, 2009) and can be adopted in certain cases to find an optimal compromise between the practical ease of data collection versus a proper geographic distribution.

Given the number and distribution of crops within the AOI, a target of 600 sample units or segments in total seems appropriate with one or two selected within a 20 x 20 km grid cell. The regular grid is preferred to the use of administrative regions as described by Defourny et al. (2019), because the size can vary considerably between counties. Multiple sample units are randomly selected for each grid cell in sequence. The selected sampling units sampled in this way are referred to as Replicate 1, Replicate 2 etc. Such an approach is well documented and was notably applied as part of the Monitoring Agriculture with Remote Sensing (MARS) programme from the European Commission (EC). A complete background is available in Gallego (1995) and Taylor et al. (1997). The advantage of using replicates is that a different sampling fraction can be applied to cropland and non-cropland strata. In addition, this approach can mitigate some of the accessibility issues by providing additional alternative sample units locations.

With this approach **588** segments have been selected for the Kenyan AOI.

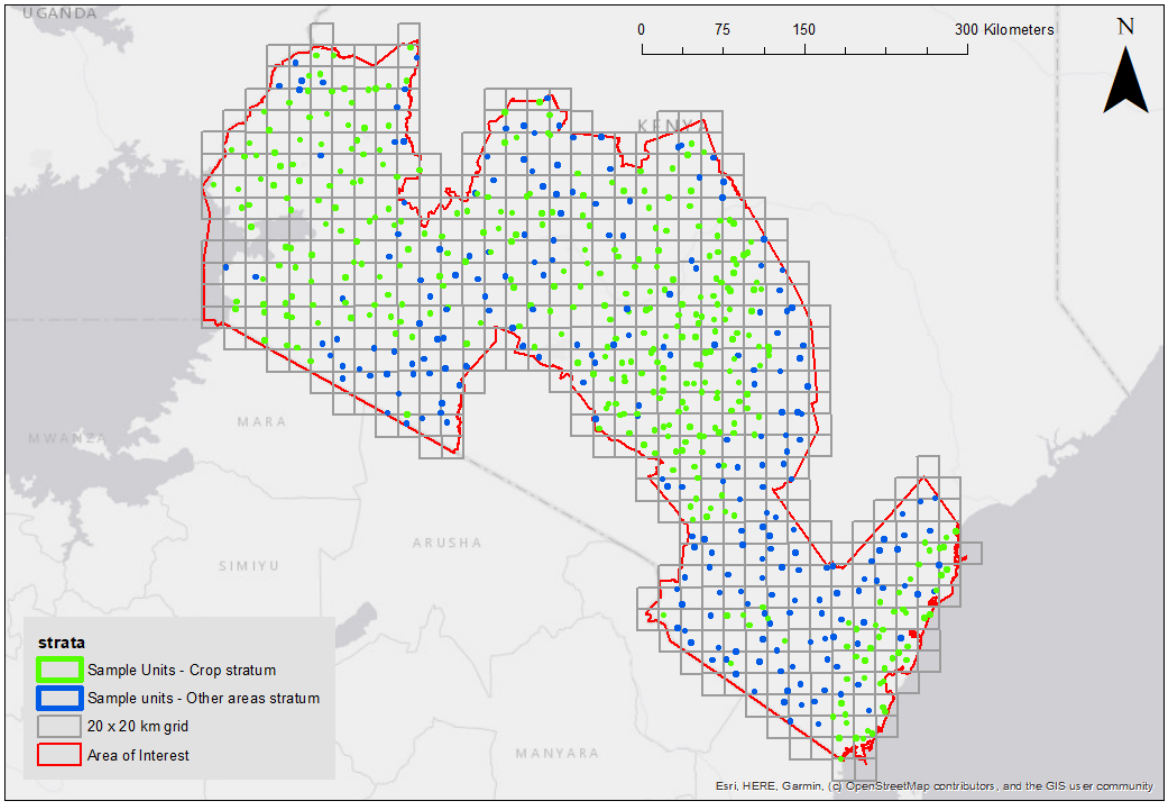


Figure 3: Spatial distribution of the sample units per aggregated stratum (data provided to JRC as shapefile)

3 Summary of data used

The Figure 4 below shows the extended AOI for Kenya, overlaid with the Sentinel-2 tile-based grid and the fieldwork (500x500m) square segments.

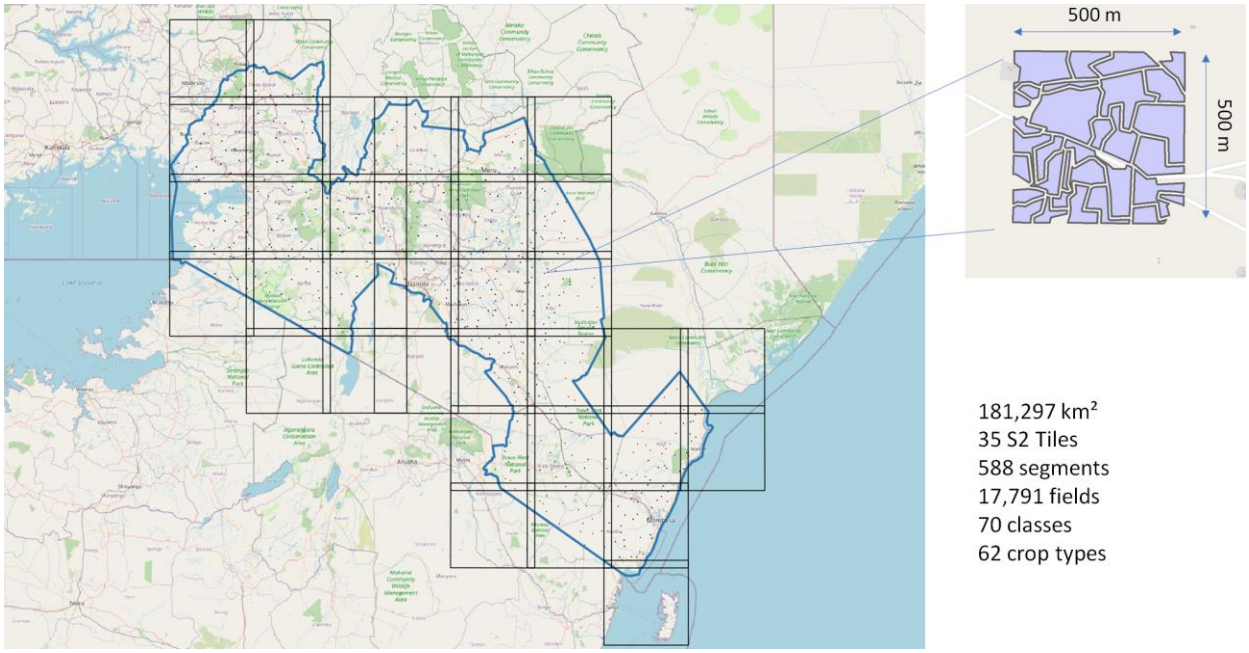


Figure 4. Kenya AOI overlaid with the S2 tile-based grid and the fieldwork segments

3.1 Satellite data – Sentinel-2

In total, 2,274 Sentinel-2A & B Level-2A images have been acquired covering 35 tiles between 01-02-2023 and 31-08-2023. The Table 1 lists the S2 data used per S2 tile ID.

Table 1. S2 tiles covering the AOI for Kenya

Tile ID	First Date	Last Date	Number of Images
36MXD	05/02/2023	29/08/2023	42
36MXE	05/02/2023	29/08/2023	41
36MYC	02/02/2023	31/08/2023	83
36MYD	05/02/2023	29/08/2023	42
36MYE	05/02/2023	29/08/2023	41
36MZC	02/02/2023	31/08/2023	84
36MZD	02/02/2023	31/08/2023	84
36MZE	02/02/2023	31/08/2023	83
36NXF	03/02/2023	29/08/2023	83
36NXG	03/02/2023	29/08/2023	80

Tile ID	First Date	Last Date	Number of Images
36NYF	05/02/2023	29/08/2023	42
36NYG	05/02/2023	29/08/2023	41
36NZF	02/02/2023	31/08/2023	84
37MBT	02/02/2023	31/08/2023	42
37MBU	02/02/2023	31/08/2023	42
37MBV	02/02/2023	31/08/2023	83
37MCR	02/02/2023	31/08/2023	85
37MCS	02/02/2023	31/08/2023	84
37MCT	02/02/2023	31/08/2023	84
37MCU	02/02/2023	31/08/2023	42
37MCV	02/02/2023	31/08/2023	42
37MDR	04/02/2023	28/08/2023	42
37MDS	02/02/2023	31/08/2023	84
37MDT	02/02/2023	31/08/2023	86
37MDU	02/02/2023	31/08/2023	85
37MDV	02/02/2023	31/08/2023	84
37MEQ	04/02/2023	28/08/2023	42
37MER	04/02/2023	28/08/2023	42
37MES	04/02/2023	28/08/2023	43
37MET	04/02/2023	28/08/2023	44
37MFS	01/02/2023	30/08/2023	89
37MFT	01/02/2023	30/08/2023	89
37NBA	02/02/2023	31/08/2023	84
37NCA	02/02/2023	31/08/2023	42
37NDA	02/02/2023	31/08/2023	84

3.2 Fieldwork data

Besides being an autonomous deliverable, the fieldwork data is also used as input into the classification procedure as well as for the validation of the results. To maximise the use of the field data in the classification workflow, the following processing steps are undertaken:

1. Assign point data (actual fieldwork) to pre-digitized polygons;
2. Apply a negative buffer of 5m to allow removal of boundary effects between landcover types;
3. Deletion of polygons smaller than 0.1 ha;
4. Splitting of data between training (75%) and validation (25%) sets;
5. Manual quality check of all training/validation polygons.

In the following, additional details regarding the five steps above are provided.

1) Data on crops and other landcover classes have been acquired in the field on the basis of pre-digitized 500x500m segments (using a combination of the most recent available Very High Resolution (VHR) imagery from Google Earth/Bing Maps, Yandex, Planet and Sentinel-2 imagery from the current season). Points have been gathered for most of the digitised polygons. Some polygons in the 500X500m segments were excluded due to the total absence of crops. For all other segments, some polygons within the segments have not been visited in the field for the safety of the enumerators (e.g. due to hostile land owners), or due to inaccessibility (e.g. flooding). To create an input for classification, point data are assigned to the polygons. In the case where no point is recorded, the land cover class recorded during the first digitising of the polygon prior to the field campaign is assigned. The polygons labelled “cropland” which are not surveyed are excluded from the fieldwork dataset since the crop type can’t be assigned. In other word, these polygons are excluded from the training dataset for the crop type mapping, from the validation and the area estimates not to bias statistics.

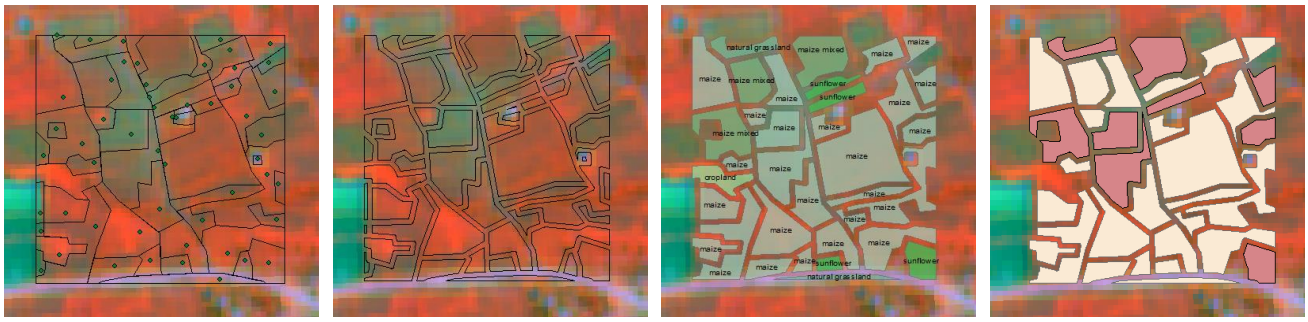
2) A negative buffer of 5 meters is applied to eliminate, or at least minimize, the boundary effects between different classes that will negatively impact the purity of training sample spectral signatures. Consequently, polygons are always separated by 10 meters, which corresponds to the size of 1 Sentinel-2 pixel.

3) The acreage of each buffered polygon is calculated and all polygons smaller than 0.1 ha are deleted. Based on the past experiences, polygons below 0.1 ha are considered spectrally heterogenous and are not deemed fit to serve as input into training samples for classification. Nevertheless, this change is the only deviation from the feasibility study report and the MMU for the classification output is still set to 0.04 ha as required.

4) All the resulting polygons have been visually checked (and manually edited where appropriate) to correct obvious thematic errors. This is done by transparent overlay of all labelled polygons per 500x500m segment over a mid-season cloud-free Sentinel-2 L3 image. Obvious labelling errors like “forest” for a cropland parcel is quickly spotted and corrected.

5) The resulting dataset from step 1 to 4 is then split into two separate sets to be used for training and validation. 75% of the dataset is used to train the classification while the remaining 25% is used for validation of the classification results. There is no overlap between the training and validation sets to ensure complete independency of the datasets. Splitting is done at a Sentinel-2 tile level to ensure a good representativity of the samples per scene. The classification workflow is applied per S2-based block as shown in Figure 4.

The Figure 5 shows for a single segment each of the above-mentioned processing steps using a Sentinel-2A L3A image from 15-04-2021 as a background.



Fieldwork points overlaid on digitized polygons

Buffered features, using inside buffer of -5m.

Removal of features < MMU (0.1 ha)

Split between training (yellow) & validation (red)

Figure 5. Preparation of fieldwork data for training and validation.

Resulting from all the described processing steps, 8,159 polygons, covering approximately 100 km² are available for the classification process. 25% are used for training and 75% for validation. In total 70 individual classes are distinguished, of which 62 individual crop types. The figures below show a few examples from the fieldwork campaign.

Figure 6. Field with maize in monoculture

Figure 7. Field with maize and cabbage in mixed cropping

End-of-season mapping methodology deviations:

For the end-of-season, there were no deviations from what was described in the methodology presented above and what was also done for the previous in-season mapping.

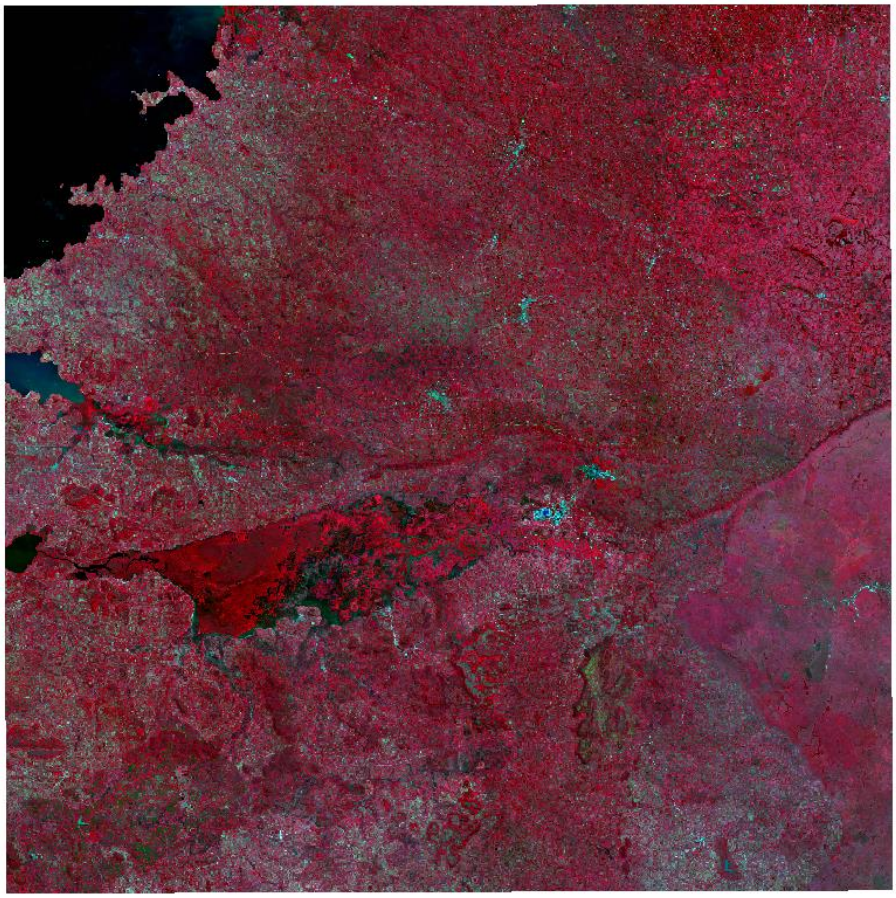
4 Workflow

4.1 Pre-processing

For both the optical and radar data, the specific pre-processing steps are summarised below.

Sentinel-2

Based on the Sentinel-2 L2A data, we reprocessed the native cloud masks using S2cloudless[4] and Fmask[5] algorithms for detailed removal of clouds and cloud shadows (see Annex IV). Monthly syntheses are then processed using the WASP algorithm (open-source solution developed by CNES[6]). For each pixel and each band (10 and 20m bands), the WASP algorithm computes the monthly synthesis using a weighted average of the cloud free surface reflectance's gathered during a synthesis period of 91 days. Cloud-free pixels as close as possible to the "centre-date" are used to build a cloud-free image. The Figure 8 shows an example for tile 36MXD, with a centre-date of 15-05-2023. For this synthesis, the algorithm considers all images +/- 45 days from the centre date and takes the cloud-free pixel closest to it.



4 https://pypi.org/project/s2cloudless/

5 jttps://doi.org/10.1016/j.rse.2014.12.014

6 https://doi.org/10.5281/zenodo.1401360

Figure 8: Sentinel-2 monthly synthesis composite, 15/05/2023, tile 36MXD.

Based on these monthly synthesis, four spectral indices are computed: the Weighted Difference Vegetation Index (WDVI[7]), Normalized Difference Vegetation Index (NDVI), Normalized Difference Water Index (NDWI), and Brightness Index (BI). See the index database[8] for the formulas and corresponding Sentinel-2 sensor bands. All bands and the indices are used as input in the classification algorithm described below.

Landsat-8

The use of the Landsat-8 dataset was not considered relevant since the L3A monthly synthesis images using Sentinel-2 were successfully generated. Moreover, the coarse spatial resolution of the Landsat-8 data (30m) was considered not suitable in case of Kenya when reviewing the size of the agricultural fields and considering the minimum mapping unit of 0.04 hectares.

End-of-season mapping methodology deviations:

For the end-of-season, there were no deviations from the proposed workflow in terms of pre-processing, which was also carried out in a similar way during the previous in-season mapping. For more information on pre-processing see Annex IV and Annex V.

4.2 Classification

Crop Type – Various classification algorithms were tested during the course of the project, including supervised (maximum likelihood) classification, TempCCN and Random Forest (RF). A full description of the classification algorithms envisaged by the consortium is given in Annex VI. It was decided to use the RF classification method for the Kenyan long rains season mapping 2023, as it consistently achieves the highest accuracy of the three methods. The algorithm is characterized by relatively simple parameterization and a good computation efficiency. Based on monthly synthesis Sentinel-2 images (L3A), precomputed features and ground truth from fieldwork (75% for training, 25% for validation), the RF classifier has been applied on all the tiles to produce the crop type map. The initial classification output contains 49 classes (of which 42 crop types). The

[7] https://www.sciencedirect.com/science/article/abs/pii/092427169190005G

[8] https://www.indexdatabase.de/db/is.php?sensor_id=96

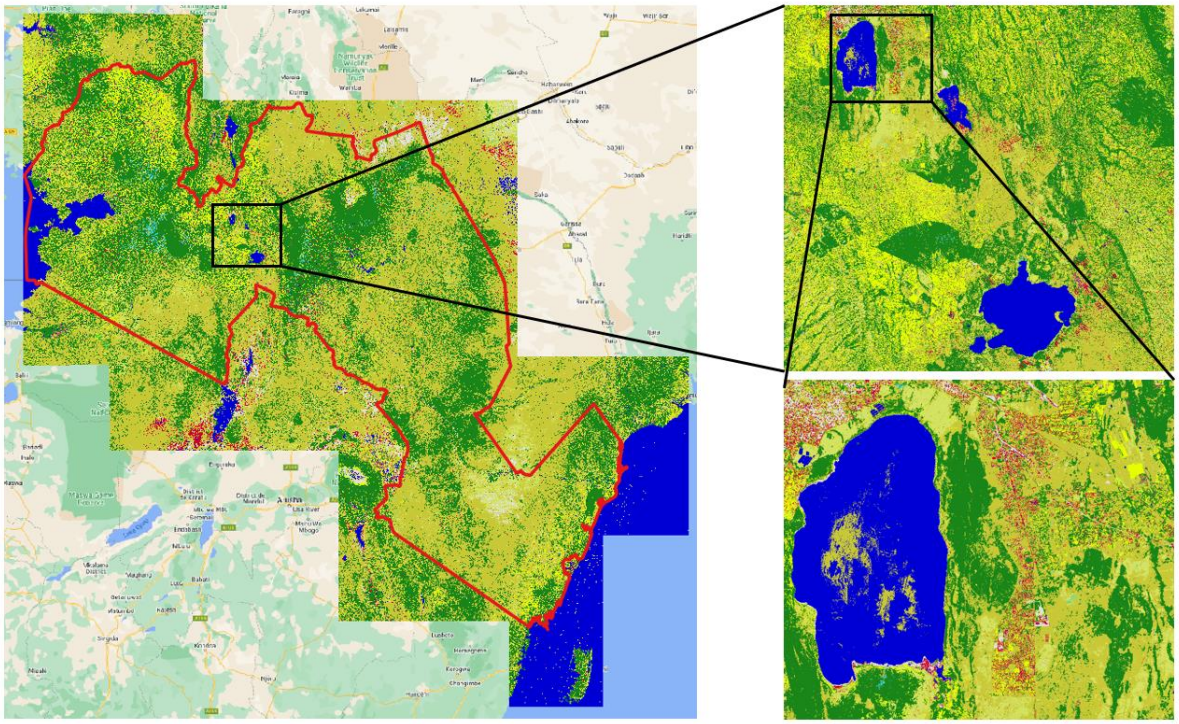


Figure 9 shows the result of the raw classification output, before post-processing.

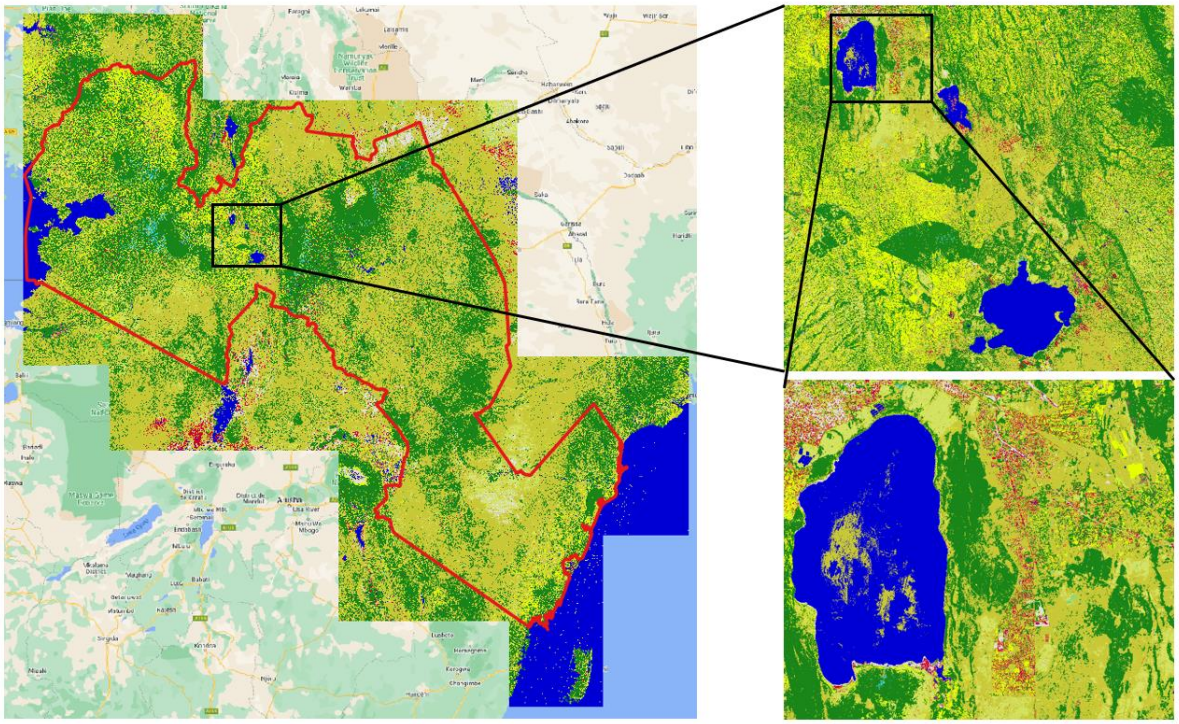


Figure 9. Raw classification output end-of-season crop type map Kenya

Crop Mask – For the crop mask, the aggregated results from the S2-derived crop type map have been used. The rule to produce the current end-of-season crop mask is as follows: If the results of the classification is

one of the 42 individual crop types or one of the mixed cropping classes it is considered *Crop*. Otherwise it is considered *Other landcover* if it is in the predicted classes of forest, natural shrubs, natural grassland, bare, urban, aquatic vegetation, water or wetlands.

The nomenclature for the Crop Mask can be found in the Table 2.

Table 2. Nomenclature for Crop Mask

Code	Class	Description
1	Crops	All monoculture and mixed cropping
2	Other landcover	Forest, water, natural shrubs, natural grassland, urban, bare, aquatic vegetation and wetlands

Post-processing of the classification results has been carried out by merging and clipping all tiles into a seamless mosaic covering the entire AOI for both Crop Type and Crop Mask. The 4 classes from the raw crop type classification are merged into 10 final classes for the final map, including the 8 largest individual crop types according to fieldwork statistics and the 5 main crops as defined by the country contact. The Table 3 lists the final classes for the Crop Type map and number coding as found in the final GeoTiff files (D3.1_Kenya_CropType_EndOfSeason_LongRains_2023.tif and D3.2_ Kenya_CropMask_EndOfSeason_LongRains_2023.tif). The nomenclature can be viewed in a GIS environment, ArcGIS or QGIS, using the accompanying *.lyr or *.qml legend layers provided with the above-mentioned GeoTiff files.

Table 3. Nomenclature for Crop Type map

Code	Class	Description
1	Maize	Including mixed cropping with maize as dominant crop
4	Beans	Including French beans, black beans and mixed cropping with beans as dominant crop
11	Sorghum	Including mixed cropping with sorghum as dominant crop
13	Green grams	Including mixed cropping with green grams as dominant crop
14	Wheat	Including mixed cropping with wheat as dominant crop
19	Sugarcane	Including mixed cropping with sugarcane as dominant crop
6	Tea	Including mixed cropping with tea as dominant crop
12	Peas	Including cow peas, chick peas, green peas, pigeon peas and mixed cropping with peas dominant crop
9	other crops	All other monoculture crops, other mixed cropping and field preparations.
10	other landcover	Forest, water, natural shrubs, natural grassland, urban, bare and wetlands.

A mask (shapefile) for all non-agricultural areas is produced from ancillary public data sources including protected area, national parks, wetland areas, open water, urban area boundaries, roads, forests and rangelands. This mask is used to recode erroneous cropland classes to other landcover, as no agriculture is (legally) supposed to be present in these areas. However, agricultural encroachment may sometimes take place in e.g. protected areas and they were preserved in the final map by a detailed visual check using recent

Sentinel-2 satellite data for verification. As a final step a sieve operation has been applied whereby all pixel clusters of 4 pixels and below (0.04 ha = approximate MMU for S2) are recoded to the majority surrounding class. All maps are presented in UTM, zone 37 South.

End-of-season mapping methodology deviations:

Compared to the proposed methodology presented in the feasibility study report (D1.1 – section 7.2.3) and what was done during the previous in-season mapping, there are no deviations to report for the end-of-season. The following are the post-processing steps to be applied as presented in the feasibility study report:

- *"Recoding: all individual classes from the raw crop type classification (usually around 45) are recoded to 10 final classes, of which 8 are major single crop type classes (tbd with country contact), an "other crops" class containing all other crop types, and an "other landcover" class containing all other non-crop classes (like water, forest, grassland, urban, shrubland). The crop mask contains 2 classes: "crops" and "no crops".*
- *Sieve filtering of pixel-based classifications: Sieve filtering recodes (clusters of) pixels to the majority surrounding class value, according to a pre-defined sieve size. It is an automated step that will be executed in QGIS.*
- *Masking of non-crop areas with ancillary data: A shapefile with non-cropped areas will be prepared for the AOI. Input layers used for Kenya consist of publicly available vector layers of protected areas, wetlands, forest and urban areas. All non-crop classes are combined into a single layer, and will then be manually reviewed in a GIS by superimposing the vector layer on a composite of recent Sentinel-2 data. Cropped areas in e.g. protected areas and forested areas (due to encroachment) are deleted from the non-crop layer. The crop type map and crop mask can be masked with the resulting layer so that mis-classified crop classes are minimised.*
- *A manual check of the classification result is done by panning through the complete mapping area in high resolution linked to a window with the corresponding satellite mosaic of a mid-season date. Some anomalies can be spotted easily (e.g. residual mis-classified landcover pixels in open water or protected areas) and can be manually recoded."*

4.3 Map production

Both the Crop Type map and Crop Mask are presented in A0 printable PDF maps with layout including legend, north arrow, metadata, grid (UTM 37, South), relevant client and contractor logo's and scale bar. The maps are presented on 1:1.000.000 scale, the largest possible scale to fit the entire AOI on A0 format. The figures below show the end-of-season Crop Mask and Crop Type map for Kenya for the long rains season.

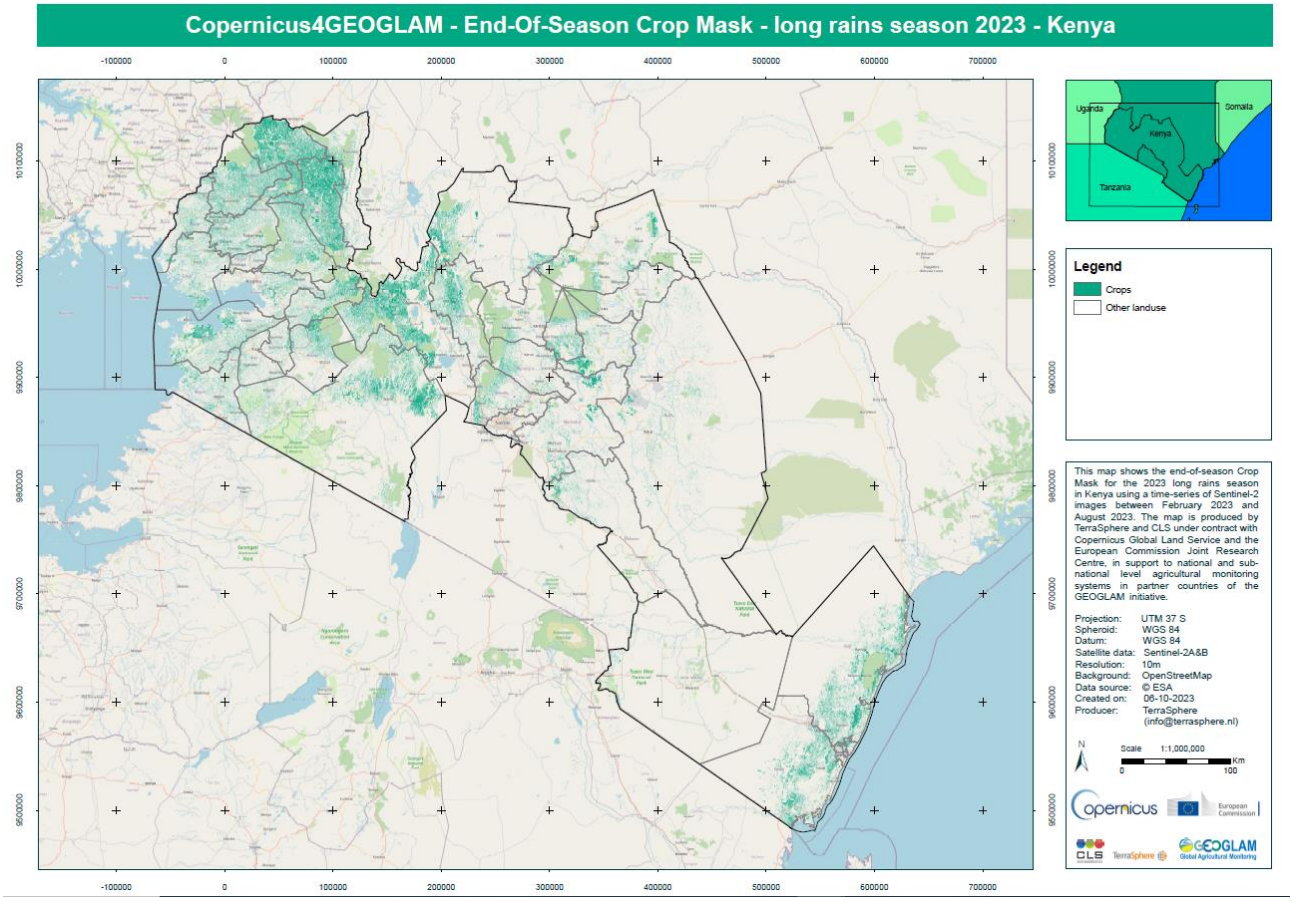


Figure 10. End-of-season Crop Mask for the long rains season 2023 in Kenya

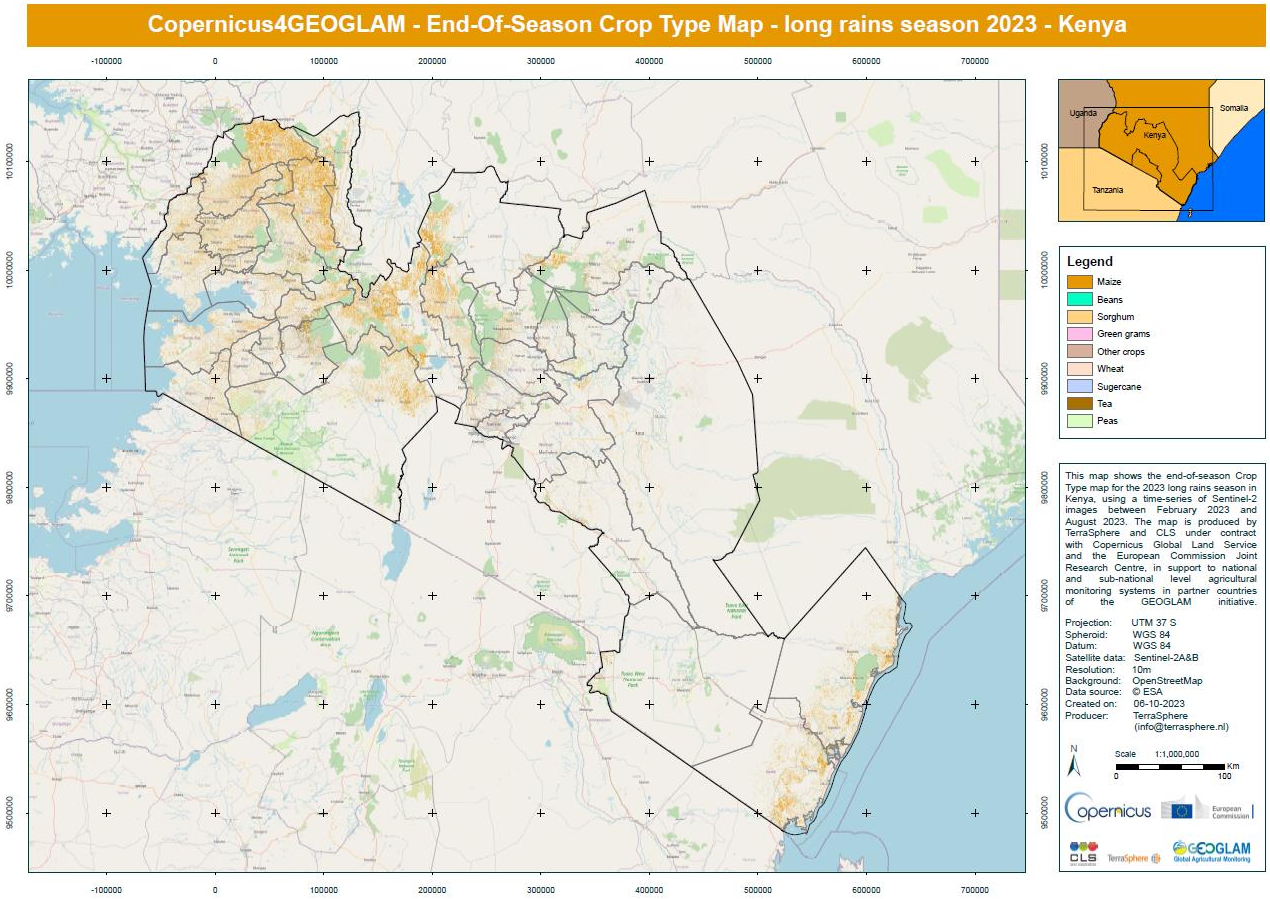


Figure 11. End-of-season Crop Type map for the long rains season 2023 in Kenya

End-of-season mapping methodology deviations:

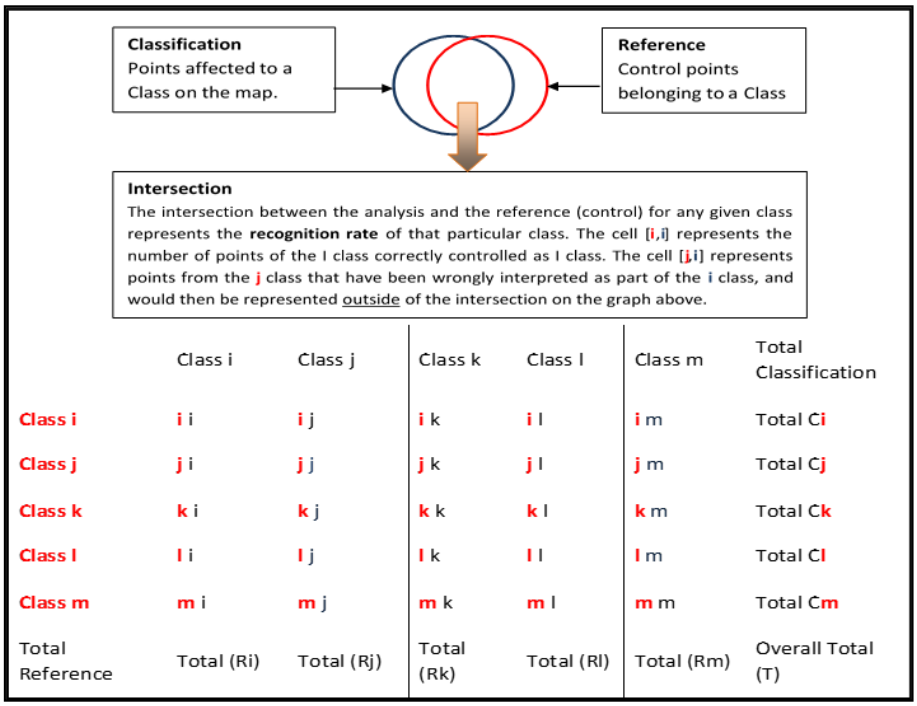
The end products are delivered in exactly the same way as the in-season mapping and carried out according to the proposed methodology and presented in the feasibility study (D1.1 – Section 7.2.4), i.e.:

- *"The Crop Type map & Crop Mask are presented in A0 printable PDF map with layout including legend, north arrow, metadata, grid (UTM 36, North), relevant client and contractor logo's and scale bar. The maps are presented on 1:600.000 scale, the largest possible scale to fit the entire AOI on A0 format."*

4.4 Validation

For both the Crop Mask and Crop Type map, 25% of processed fieldwork data (that is not used for training) is used for validation. Confusion matrices are produced and F1 score per class have been calculated and can be found in the figures below. There was no need to apply correction factors because an equal sampling intensity was applied to each stratum. The validation procedure was carried out as described in the feasibility study and is presented below.

Thematic accuracy is assessed based on the construction of a confusion or error matrix. The construction of the error matrix is described as illustrated in Figure 12 for 5 thematic classes.



	Class i	Class j	Class k	Class l	Class m	Total Classification
Class i	i i	i j	i k	i l	i m	Total Ci
Class j	j i	j j	j k	j l	j m	Total Cj
Class k	k i	k j	k k	k l	k m	Total Ck
Class l	l i	l j	l k	l l	l m	Total Cl
Class m	m i	m j	m k	m l	m m	Total Cm
Total Reference	Total (Ri)	Total (Rj)	Total (Rk)	Total (Rl)	Total (Rm)	Overall Total (T)

Figure 12: Confusion Matrix for Accuracy Assessment of thematic map product

Then, the following accuracy metrics can be calculated from the confusion matrix:

- The **Overall Accuracy** or **Recognition Rate** is measured by the sum of the diagonal of the Confusion Matrix divided by the total number of controlled points: $OA\ or\ R_r = \sum_{\alpha=i}^{m}(\alpha\alpha)/\mathrm{T}$. The **Recognition Rate** or **Overall Accuracy** assesses the overall agreement between the classified and reference data set. However, for single class themes, it does not necessarily provide a realistic assessment of the quality of the map produced because there can be substantial unbalance between omission and commission errors.
- Therefore, the row and column totals and the diagonal of the Matrix are used to assess two types of accuracy, the User's and Producer's Accuracy:
    - **Producer Accuracy** for the α class = $\alpha\alpha/\mathrm{C}\alpha$ is a measure of **omission error**. For instance, an observation has been identified as maize in the validation set, but has been classified as another class: it has been omitted from the maize class.

- **User Accuracy** for the α class = $\alpha\alpha/R\alpha$ is a measure of the **commission error** (or contamination risk): errors due to the wrong allocation of an observation to a class. For instance, an observation is classified as maize, but identified as belonging to another class during the validation process: this observation has contaminated another class.

A full description of the validation strategy is given in Annex VII.

Figure 13 and Figure 14 show that the overall accuracy for the Crop Type map and Crop Mask are respectively 89% and 90%, higher than the specifications mentioned in the feasibility study report (D1.1) (65% for both).

Crop type end-of-season mapping for Kenya (long rains)	Reference												
		Maize	Beans	Sorghum	Green grams	Wheat	Sugarcane	Tea	Peas	Other crops	Other landcover	Total	User Accuracy	F-Score
Map	Maize	**12245**	8	11		495	193	1	127	666	3954	**17700**	0,69	0,60
	Beans	35	**7**								9	**51**	0,14	0,05
	Sorghum											**0**	N/A	N/A
	Green grams				**18**							**18**	1,00	0,15
	Wheat	157				**880**	1		12	4	2	**1056**	0,83	0,64
	Sugarcane	14					**336**				279	**629**	0,53	0,29
	Tea	18						**1044**		40	28	**1130**	0,92	0,81
	Peas	101							**61**		205	**367**	0,17	0,06
	Other crops	138		18	36			18	83	**1516**	533	**2342**	0,65	0,29
	Other landcover	10586	202	126	173	330	1198	382	1321	6063	**213215**	**233596**	0,91	0,94
	Total	**23294**	**217**	**155**	**227**	**1705**	**1728**	**1445**	**1604**	**8289**	**218225**	**256889**		
	Producer accuracy	0,53	0,03	0,00	0,08	0,52	0,19	0,72	0,04	0,18	0,98			

Overall accuracy	0,89
Crop Type User accuracy	0,69
Crop Type Producer accuracy	0,42

Figure 13. Confusion matrix for end-of-season Crop Type map of the long rains season 2023

Cropland mask end-of-season mapping for Kenya (long rains)		Reference				
		Non-crop	Crop	Total	User Accuracy	F-Score
Map	Non-crop	213215	20381	233596	0,91	0,94
	Crop	5010	18283	23293	0,78	0,59
	Total	218225	38664	256889		
	Producer accuracy	0,98	0,47		Overall accuracy	0,90

Figure 14. Confusion matrix for end-of-season Crop Mask of the long rains season 2023

The crop mask for the end-of-season shows satisfying results for the user accuracy of the crop class (78%) with 22% of commission errors; meaning that the map isn't too much contaminated by erroneous crop detection. The producer accuracy is still low (47%) with 53% of omission errors. The classification has missed some cropped areas but large improvements have to be noted (+18%) from the in-season to the end-of-season with the addition of S2 images at the beginning and at the end of the season to better depict the start of the growing season and the late stage of the harvest.

The omission phenomenon is also confirmed by the confusion matrix of the crop type map, which shows low producer accuracies for most of the classes (lower than 10%); except for tea, maize and wheat with a producer accuracy of 72%, 53% and 52%. The matrix shows very satisfying user accuracies for the different crop types (except beans and peas). For example, the commission errors for the maize, wheat and tea classes are respectively 31%, 17% and 8%. As only few samples were available for class green grams, the 100% producer accuracy should be taken with caution.

Overall, the wheat, maize and tea classes tend to show satisfying results with F1-Score between 0.60 and 0.81. For other classes, results are very low with F1-Score below 0.15 (beans, green grams and peas).

The low results obtained for these classes can be explained by a number of factors, but the small size and the low number of usable fields (> 0.1 ha) associated with the training data are probably one of the most important. Most small-scale farmers in Kenya practice subsistence farming, growing crops primarily to meet the needs of their families and local communities rather than for commercial purposes. The average field size in the training dataset for e.g. beans, green grams and peas is 0.4 ha whereas the average field size for the other crops is 1.8 ha, which is 4.5 times higher. As a result, many fields are too small to be used for training.

The limited number of small polygons for these classes results in heterogeneous training data and consequently low classification accuracies.

Mixed cropping/intercropping may also have a small effect on the results. Even if the mixed cropping fields are discarded for the learning phase, the fields with a declared dominant crop (more than 50% of the parcel) are kept for the training dataset. Yet, a lot of small-scale farmers in Kenya practice mixed cropping farming. The total field size for the mixed cropping in the dataset is 1,046 km² out of 2,774 km², representing 37.7% of the cropland polygons. A total of 1,734 polygons (representing 855 km² out of 1,046 km²) in mixed cropping with a dominant crop declared are kept for the training. As a result, these fields can have a heterogeneous spectral signature, resulting in low classification accuracies.

To a lesser extent, other reasons for low results could be the quality of the field work or the inaccurate delineation of the polygons/fields.

The contribution of each reason to the results is hardly impossible to estimate and assumptions should be made with caution.

End-of-season mapping methodology deviations:

For the end-of-season, there were no deviations from the methodology described above and in the feasibility study (D1.1) and how the statistics were performed during the previous in-season mapping.

4.5 Area estimates

As described in the feasibility study report (D1.1), crop area statistics are also provided, including:

1. Direct expansion estimates: area estimates from the field data alone;
2. Pixel count: areas measured from the end-of-season map alone;
3. Regression estimators: area estimates derived from field data combined with in- season map based on linear regression.

In the following, additional details regarding the three estimates are provided.

(1) Crop area estimates can be derived directly from the field data alone using the so-called direct expansion method since the data has been collected based on a probabilistic sample. Nevertheless, the confidence interval of the estimates derived from direct expansion is relatively large. To better consider the mixed cropping practice, all the crop surveyed in the field were taking into account for the estimates:

1) contributing equally to the total area of the field if no dominant crop was declared or,
2) the dominant crop contributing to half of the total area and the other crops surveyed contributing equally to the second half of the total.

Figure 15 illustrates the change with one example.

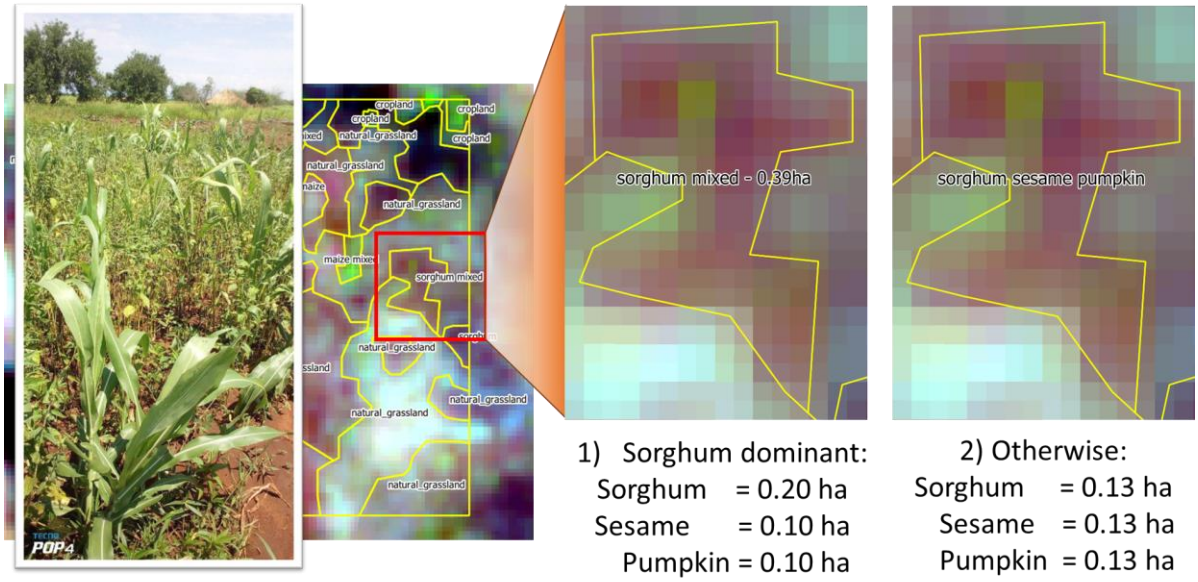


Figure 15 : Mixed cropping fields and crop area estimates (non-dominant crop study case)

(2) Crop area estimates can be derived directly from the end-of-season map alone. Areas measured from digital classification have no sampling errors because they are based on pixel counts covering the whole of the AOI but they are biased because of misclassification. These misclassifications can be omission and commission errors (respectively (omitted features and contamination of a given feature in the final map). So area estimates from digital classification are not accurate because significant numbers of pixels of the Sentinel-2 image can be mis-classified.

(3) To improve the precision of the estimates, field segment data (1) can be combined with classified satellite imagery (2). In this latter case, a Regression Estimator model can be applied which is more reliable than any other area estimation methodology as it provides both an area estimation per cover type together with an indication of its uncertainty. In brief, Regression Estimator relies on the combination of area estimates made

at the segment level for both ground data and classified satellite imagery. The observations are paired, and a regression analysis is performed. The regression estimator y_{reg} is calculated based on the following equation:

$$y_{reg} = \bar{y} + b * (\bar{p}_{pop} - \bar{p})$$

Where: $\bar{y}$ is the mean field data sample value, b is the slope of the regression line, $\bar{p}_{pop}$ is the proportion of pixels classified as the crop in the whole of the region of interest and $\bar{p}$ is the classified image mean sample value.

A full description of the area estimate bias correction is given in Annex VIII.

The improvement achieved by combining field data with satellite classification can be calculated by a metric called the relative efficiency. The relative efficiency is used to estimate the additional size of field work sample needed to achieve an equivalent improvement in accuracy of area estimates. Table 4 shows the results of the crop area estimates for Kenya. Very good relative efficiencies for Maize, Wheat, Sugarcane and Tea with figures greater than 2 are to be noticed. For Beans, Sorghum, Green grams and Peas, the relative efficiencies are lower than 2. For example for Wheat, Rice and Sugarcane, the same reduction in variance would have been achieved by increasing the size of the field survey sample by 19.5, 9.1 and 7.6. Due to time and budget constraints, it is clear that such an increase in the field campaign is not feasible. This would mean visiting more than 3,000 sample units (500x500m square segments)!

Table 4: Area estimates for the end-of-season mapping of the long rains season 2023 in Kenya

AOI Area (ha)		18 102 432,96	Maize	Beans	Sorghum	Green grams	Wheat	Sugarcane	Tea	Peas	Other crops	Other landcover
Direct Expansion	Estimate of proportion		0,10	0,01	0,00	0,01	0,01	0,01	0,01	0,02	0,03	0,82
	Variance		0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	Standard Error		0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,01
	95% Confidence Interval		0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,02
	Estimate of the class area		**1 736 202,78**	**121 905,67**	**62 120,04**	**103 966,15**	**107 555,34**	**137 291,80**	**127 604,73**	**324 035,55**	**521 877,67**	**14 859 873,22**
	Variance		10 807 395 865,30	258 994 420,79	101 691 880,77	281 462 014,48	1 152 996 317,01	1 075 918 733,45	1 174 428 415,57	1 384 903 128,26	2 169 255 317,11	22 222 922 884,00
	Standard Error		103 958,63	16 093,30	10 084,24	16 776,83	33 955,80	32 801,20	34 269,93	37 214,29	46 575,27	149 073,55
	95% Confidence Interval		203 758,91	31 542,88	19 765,11	32 882,59	66 553,37	64 290,35	67 169,07	72 940,00	91 287,52	292 184,16
Pixel count	Map (ha)		**1 249 435,98**	**294,98**	**547,27**	**47,27**	**13 484,28**	**16 029,94**	**107 113,24**	**1 074,57**	**34 271,16**	**16 680 134,28**
	Map (%)		0,07	0,00	0,00	0,00	0,00	0,00	0,01	0,00	0,00	0,92
Regression Estimator	Regression estimate		0,09	0,01	0,00	0,00	0,00	0,00	0,01	0,02	0,02	0,85
	Variance		0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	Standard Error		0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	95% Confidence Interval		0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	Regression estimate of the class area		**1 600 626,85**	**104 009,78**	**47 227,13**	**80 915,40**	**42 294,52**	**50 458,32**	**148 443,63**	**274 703,66**	**423 985,73**	**15 306 949,95**
	Variance		3 070 556 250,34	227 157 606,90	69 355 336,10	205 050 794,62	126 793 687,04	55 286 977,31	155 440 566,48	1 025 085 904,70	1 807 017 390,68	7 975 023 430,86
	Standard Error		55 412,60	15 071,75	8 327,99	14 319,59	11 260,27	7 435,52	12 467,58	32 016,96	42 509,03	89 302,99
	95% Confidence Interval		108 608,70	29 540,63	16 322,85	28 066,41	22 070,13	14 573,62	24 436,46	62 753,25	83 317,69	175 033,85
Efficiency	Regression Estimator		**3,52**	**1,14**	**1,47**	**1,37**	**9,09**	**19,46**	**7,56**	**1,35**	**1,20**	**2,79**

5 Conclusions

The overall accuracies for the long rains end-of-season Crop Type map and Crop Mask for the long rains are respectively 89% and 90%, which is higher than what was mentioned in the feasibility study (both 65%). Some classes such as wheat, maize and tea show good results (with F1-Score between 0.60 and 0.81) but for some individual crops (e.g. beans, green grams and peas), lower accuracies are reported.

Overall, the user accuracy is relatively good, meaning that there are few commission errors in the products. The two maps are not contaminated too much by erroneous crop detection. Nevertheless, the producer accuracy of the two products is lower resulting in omission errors, which means that the classification has missed quit some cropped areas. A large improvement (+18%) is noted from the in-season to the end-of-season with the addition of S2 images at the beginning and at the end of the season to better depict the start of the growing season and the late stage of the harvest.

The low results obtained for these classes can be explained by the small size and the low number of usable fields (> 0.1 ha) associated with the training data. As a result, many fields are too small to be used for training. The limited number of small polygons for these classes results in heterogeneous training data and consequently low classification accuracies.

Very good relative efficiencies for Maize, Wheat, Sugarcane and Tea with figures greater than 2 are to be noticed. For Beans, Sorghum, Green grams and Peas, the relative efficiencies are lower than 2.

Annexes

Annex I. Stratification approach

Stratification is useful in this context of crop mapping because it can reduce the amount of effort required for the field campaign substantially. By minimising the number of sample units in non-cropped areas, the field campaign can be executed more efficiently without losing precision.

In many cases, the land cover map to be assessed or validated is directly used as stratification (Lowell and Jaton, 2000). This is a good approach to estimate commission errors in a binary classification but may be weak for other cases. If we have some information that quantifies the likely proportion of errors, for example a measure of landscape complexity, it can be a more efficient stratification for all types of errors. We shall choose a higher sampling rate in more difficult areas, where both the errors and their variances are higher.

However, in this context, the field campaign and thus the stratification plan is required prior to producing the map.

Experience has shown that a too complex stratification will not bring major improvement and that there must be a clear case for stratifying the AOI. A good stratification should reduce the variability within each stratum, and considering the complexity of cropping patterns, it may sometimes be difficult to accurately describe the reality. A too complex and insufficiently accurate stratification may be counterproductive and lead to substantial variability with wider 95% confidence interval of the resulting crop area estimates as compared with simple random sampling. A clear case for stratifying is when there are marked geographical differences in the landscape between areas that are cultivated versus areas for which there is little agriculture present. A potential approach is to separate between different cropping systems based on clearly defined characteristics such as (i) Agro-climatic conditions or (ii) on the identification of irrigated versus non-irrigated land.

In cases where agro-climatic conditions or crop systems are difficult to identify based on available data, crop area/production statistics per administrative unit can be used. In this case, provided that the crop area statistics are sufficiently accurate, strata can be separated according to districts which are predominantly agricultural versus those in which agriculture is marginal. However, crop statistics at some administrative level should be used carefully as described by Sannier et al. (1998). Crop statistics and biophysical factors can be combined in a relatively complex analysis to stratify the area, but experience in Europe from the Regional Inventory programme of the MARS project in the 1990s (Taylor, 1997) showed that the more complex is the analysis put in place to stratify the area, the less efficient the stratification tends to be. In addition, the total number of strata should not be large because, even if the stratification allows to improve the efficiency of the sampling by reducing the number of sample units per stratum, the total number of sampling units may be larger if too many strata are present. Pragmatically, Cochran (1977, p. 134) recommends no more than 6-8 strata in total.

In conclusion, considering good practices from the literature and past experience, we recommend selecting no more than 6 to 8 strata. This is in line with the technical specifications, mentioning that if the AOI is not homogenous, it should be subdivided into strata from 2 to 5 strata, which are assumed to be homogeneous regarding climate and agro-ecological conditions (relief, soil, etc.) as well as agricultural practices (e.g., crop calendars).

Annex II. Field sampling strategy

Two types of protocol exist regarding the collection of field data: 1) a probability sampling protocol and 2) a "windshield" survey with collection of data along the roads. A probability sampling design is essential for map validation and area estimation (Olofsson et al., 2013; Stehman et al., 2009), but this is less critical for collecting training data for satellite image classification. So alternatively, a selection of locations for training data can be a pragmatic and suitable choice, for example collecting data along roads (Gallego, 2018). The "Windshield survey" can complement the probability sampling protocol to provide a larger sample size for the training of the classification algorithms.

In the following paragraphs, we first describe the probability sampling protocol and then the complementary windshield survey with the pros and cons of the two approaches.

Probability sampling protocol

Simple random, stratified random, clustered random and systematic designs are all examples of probability sampling designs. In simple random designs, classes covering a small portion of the population may not be adequately sampled. Clustered sampling is often used to reduce the costs of the collection of reference data but does not resolve geographic distribution problems. Stratified approaches overcome this drawback; therefore, the stratified random sampling of points is one of the most common approaches to assess map thematic accuracy but may not be the best option for assessing areas. Stratified systematic with random origin has the advantage of enhancing traceability and is a better solution for assessing areas. The main limitation of systematic sampling is that there are no unbiased estimators of the variance, and the simple random sampling variance is often used which would slightly overestimate the true value (Bellhouse, 1988).

The provision of accurate area estimates requires some form of unbiased independent assessment of the map quality. The so-called "pixel counting" on classified images cannot be considered a sound approach for crop area estimation mainly because remotely sensed based crop and more generally land cover classification suffer from classification errors that affect each crop type or land cover class differently. This means that a specific crop or land cover type can be over-classified while another can be under-classified, thus over- or under-estimating the area of one versus the other. This was already pointed out by many in the scientific literature as early as Hay (1988) and Czaplewski (1992) and more recently by Olofsson et al. (2014). However, to be able to assess such bias in crop type maps, there is a requirement to use a probabilistic sample to ensure that the estimates can be produced directly from the probabilistic sample or from the combination of the probabilistic sample with the classified image.

A probabilistic sample is one in which the probability of inclusion of a sample can be determined. It can typically be achieved through a random or systematic approach combined or not with a stratification as shown by the early implementation of satellite-based crop area estimation systems in the USA (Allen, 1990) and in Europe as part of the MARS project in the late 1980s and early 1990s by (Taylor, 1997). In both crop area estimation activities, the existing field survey data was combined with the satellite-based classification maps through a so-called regression estimator to provide unbiased crop area estimates.

In summary, a probabilistic sample approach is needed for unbiased crop area estimate and will by default applied in the present service. Among the possible probabilistic approaches, we recommend a stratified systematic random sampling which combines benefits from both stratified systematic and random protocols.

"Windshield survey"

Current machine learning algorithm will perform best with large amount of training data. Collecting field data in a "windshield survey" along the road network as suggested by (Defourny et al., 2019) can provide large amount of field data with reduced effort. However, it does not constitute a probability sample because sample units are defined a priori (i.e. by the road network) and there is no way to calculate their probability

to be included in the sample. Therefore, this approach does not seem appropriate for providing reliable area estimates. In fact, the JECAM guidelines (*JECAM, Guidelines for cropland and crop type definition and field data collection*, 2018) recognises this limitation by suggesting that the windshield survey is insufficient: "It is however recommended that this sampling strategy be complemented by regular additional transect (set 2) using secondary roads and tracks to reduce the spatial bias brought about by roadside sampling" (*JECAM, Guidelines for cropland and crop type definition and field data collection*, 2018).

Evidences of bias resulting from this type of survey can be found in a study performed by Gallego (2018). Based on data collected with a probabilistic approach (LUCAS - Land Use and Coverage Area frame Survey in Europe), the author indicates there could be a substantial bias when considering LUCAS points only in the close vicinity of roads (less than 100m). Results show that for some main crops such as wheat this may not be too problematic, but there could be more substantial effects for less represented crops such as barley and maize and particularly for Olive trees and vineyards as shown in Table 5. Yet, this study was carried out in Europe for which the road network is particularly dense, and therefore the bias is expected to be even larger in developing countries where road networks can be much sparser.

Table 5: Weighted estimation of the proportion of major crops from all LUCAS points and from LUCAS points within 100 m of a road. Gallego (2018) and Gallego, personal communication

	All LUCAS points	*Points within 100 m of a road*	*% bias*
Wheat	24.88	25.24	1.4
Barley	14.63	14.28	-2.4
Maize	12.31	11.94	-3.0
Root crops	3.62	3.65	0.8
Sunflower	2.50	2.53	1.2
rapeseed	6.18	6.24	1.0
fodder & temp grass	7.07	6.70	-5.2
olive trees	5.83	6.04	3.6
vineyards	3.75	3.97	5.9

In conclusion, the "windshield survey" is not recommended for unbiased crop area estimates but the approach is perfectly suitable for the collection of additional training data, since the collection of field data for image classification does not necessarily need to be probabilistic.

Annex III. Sample size per stratum

When using stratified sampling, the main goal is to maximise the efficiency of the stratification by optimising the sample allocation per strata with a view to (1) minimise the overall number of samples necessary (by reducing the number of samples for strata with little agriculture present) and/or (2) minimise the uncertainty of the resulting crop area estimates. A simple way is the use of equal allocation, but this is usually not very efficient. Proportional allocation (to the fraction of the total area covered by a given crop) is an option, but it will give disappointing results for classes considered important but covering a small proportion, e.g. crop types covering a small area present only in very specific strata.

It is possible to estimate a suitable sample size for each stratum based on the expected acceptable error rate which can be referred to the error rate requested by the technical specifications and or previous test results. For example, it may be required that the area of wheat at the scale of the AOI should be estimated with an uncertainty that does not exceed 10% at 95% confidence interval. The standard error of the error rate can be calculated as follows:

$$\sigma_h = \sqrt{\frac{p_h(1-p_h)}{n_h}}$$

where n_h is the sample size for stratum h and p_h is the expected error rate.

This can be reworked to express the sample size n_h as a function of p_h and desired standard error σ_h:

$$n_h = \frac{p_h(1-p_h)}{\sigma_h^2}$$

From Figure 16, it can be seen that for an expected 50% error rate, within a stratum, 100 samples would be required to guarantee a standard error of 5%, whereas the number of samples would need to be increased by a factor of four if the accepted standard deviation is divided by a factor of 2. On the other hand, if the expected error rate is 15%, only 51 samples would be necessary with a 5% standard error. A similar approach was adopted to determine the sample size for assessing the accuracy of CORINE Land Cover (CLC) 2006 and CLC 2000-2006 changes (Büttner et al., 2012).

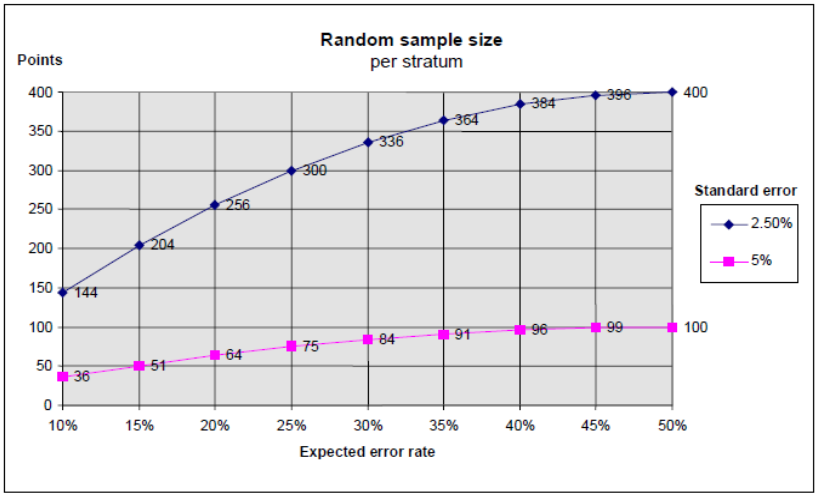


Figure 16: Number of sample points as a function of the expected error rate for two accepted standard error values (after Wack et al., 2012)

If we have a high priority for a class, the Neyman allocation rule is a better alternative (Cochran, 1977):

$$n_h = n * (N_h * \sigma_h) / [\Sigma (N_i * \sigma_i)]$$

where n_h is the sample size for stratum *h*, *n* is the total sample size, N_h is the population size for stratum *h*, σ_h is the standard error of the error rate of stratum *h, i* extends to all strata within the AOI.

According to (Stehman, 2012), Neyman optimal allocation should be preferred for estimating area as well as overall accuracy, whereas equal allocation is effective for estimating user's accuracy. For our purpose, it would be difficult to apply Neyman allocation for each crop type, but it can be used for generically for crop areas optimising allocation where agriculture represents a large proportion of the area.

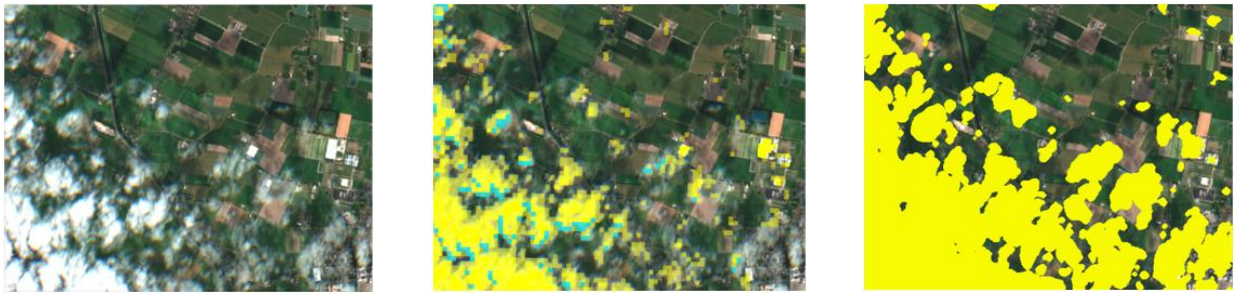
In practice (Särndal et al., 1992, pp. 267 and 407) recommend a minimum within-stratum sample sizes of 10-20; (Cochran, 1977, p. 134) recommends minimum within-stratum sample sizes of 20; and for temperate forest inventories (Westfall et al., 2011) recommend within-stratum sample sizes of at least 20.

Annex IV. High Spatial Resolution Optical data: Sentinel-2

Copernicus Sentinel - data are now available since 2015 for Sentinel 2A and since 2017 for Sentinel-2B. They provide multispectral data, 13 bands with a spatial resolution of 10, 20 and 60 meters with a large swath. The high-revisit frequency of acquisitions (maximum 5 days) allows to build dense times series of high-resolution optical data. However, the impact of clouds (and cloud shadows) can decrease drastically its high frequency and requires the use of performant cloud detection algorithms.

Times series of Sentinel-2 data are used to compute pixel-based classification at 10 meters. To perform these tasks, the pre-processing of the Sentinel-2 data must ensure a high confidence cloud (and cloud shadows) detection and a high geometric temporal alignment. The pre-processing step will create a set of Analysis Ready Data (ARD) used for the generation of crop type and mask products.

For Sentinel-2, we consider currently only L2A provided by ESA (and generated by Sen2Cor) with surface reflectance and scene identification (cloud and cloud shadow masking mainly). As we know that the quality of the scene identification and mainly the cloud and cloud shadow identification is not optimal (see Figure 17), we propose the following alternative scenario: if the quality of cloud identification has a strong impact on the quality of the segmentation and the pixel identification, we will perform the L2A generation with the Multi-sensor Atmospheric Correction and Cloud Screening (MACCS)-Atmospheric & Topographic Correction (ATCOR) Joint Algorithm (MAJA).



(a) (b) (c)

Figure 17: Cloud cover and native cloud masking within a Sentinel-2 image: (a) clouded S2 image; (b) native (sen2core) cloud (yellow) and cloud shadows(blue) masks, and (c) S2cloudless cloud masking result.

The native scene identification mask delivered with the sentinel-2 data has some drawbacks; 1) coarse mask with pixels of 60 meters, 2) not optimal cloud detection and 3) not precise matching of cloud shadows. An alternative masking procedure can be the following: using the S2cloudless[9] algorithm in combination with an adapted F-mask algorithm. When using the algorithm, the consortium found that in terms of speed and ability to detect (even the smallest) clouds, the S2cloudless algorithm yielded excellent results. Also, a test performed by the S2cloudless developers indicated good results when compared to other algorithms like Sen2cor, MAJA and F-mask[10] (see Figure 17 (c)).

The drawback of the S2cloudless algorithm is that it does not provide masking for other disturbances, mainly being cloud shadows. Our adapted F-mask algorithm can be fed by the result of the S2cloudless algorithm and provide the other masks. Our F-mask algorithm can predict the cloud shadows based on cloud clumps, the sun and satellite angles and the presence of darker pixels. Together a reliable mask can be generated.

[9] https://github.com/sentinel-hub/sentinel2-cloud-detector

[10] https://medium.com/sentinel-hub/improving-cloud-detection-with-machine-learning-c09dc5d7cf13

The second challenge to tackle with pre-processing is to reach high geometric temporal alignment between scenes. It is known that small shifts between S2 scenes occur (especially compared between S2-A and S2-B), which can be non-trivial especially for smaller fields within range of the MMU. In order to ensure perfect alignment, shifts can be applied to (parts of the) scene. It uses specific bands from the scene and a template to match features. The calculations are performed on a Google Graphics Processing Unit (GPU) box and yields a point-grid, whereby each point indicates the shifts applicable for that area between S2-A and S2-B and when available against better reference data (see Figure 18). Shifts can be non-consistent throughout the scene, so they are applied in a later stage locally in a dedicated non-linear warping procedure.

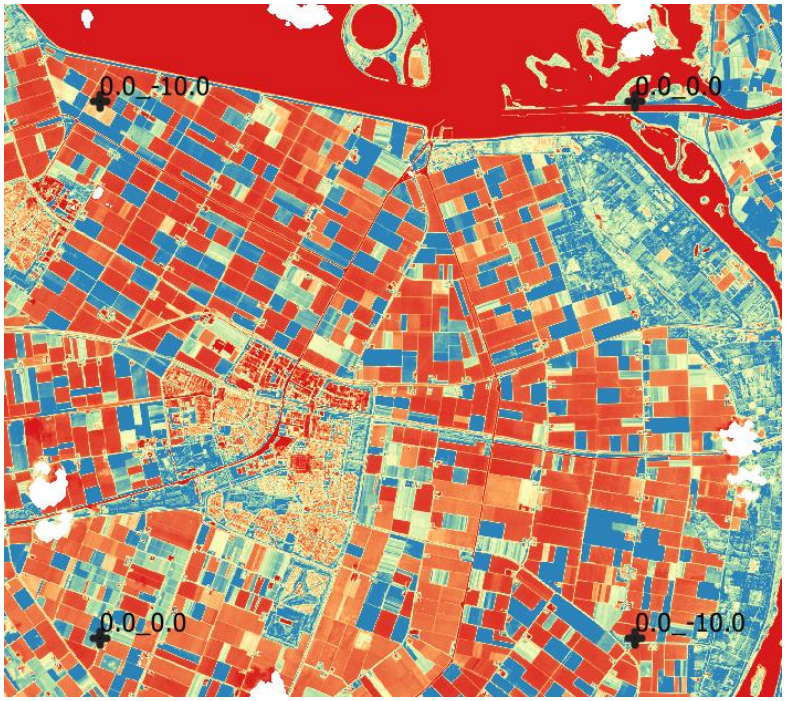


Figure 18: Example of XY shifts in meters in a grid for a Sentinel-2 scene

From the Sentinel-2 L2A product, we will keep for the following processing only the 10- and 20-meters bands (10 bands). In order to increase the usability of the Sentinel-2 data, we will resample the 20 metres bands to 10 metres with a zooming algorithm based on Fourier transform using a super-resolution algorithm provided by (Brodu, 2017) and its application to Sentinel-2 images. A reference open-source implementation already exists and is available in SNAP. After this step, we provide Sentinel-2 data with the 10 bands at 10 metres.

All the quality masks are combined and processed to generate a Usable Data Mask (UDM) which identifies the clouds, the no data, snow, the pixel quality (defective, saturated, etc.).

For a schematic overview of the points mentioned above, see Figure 19.

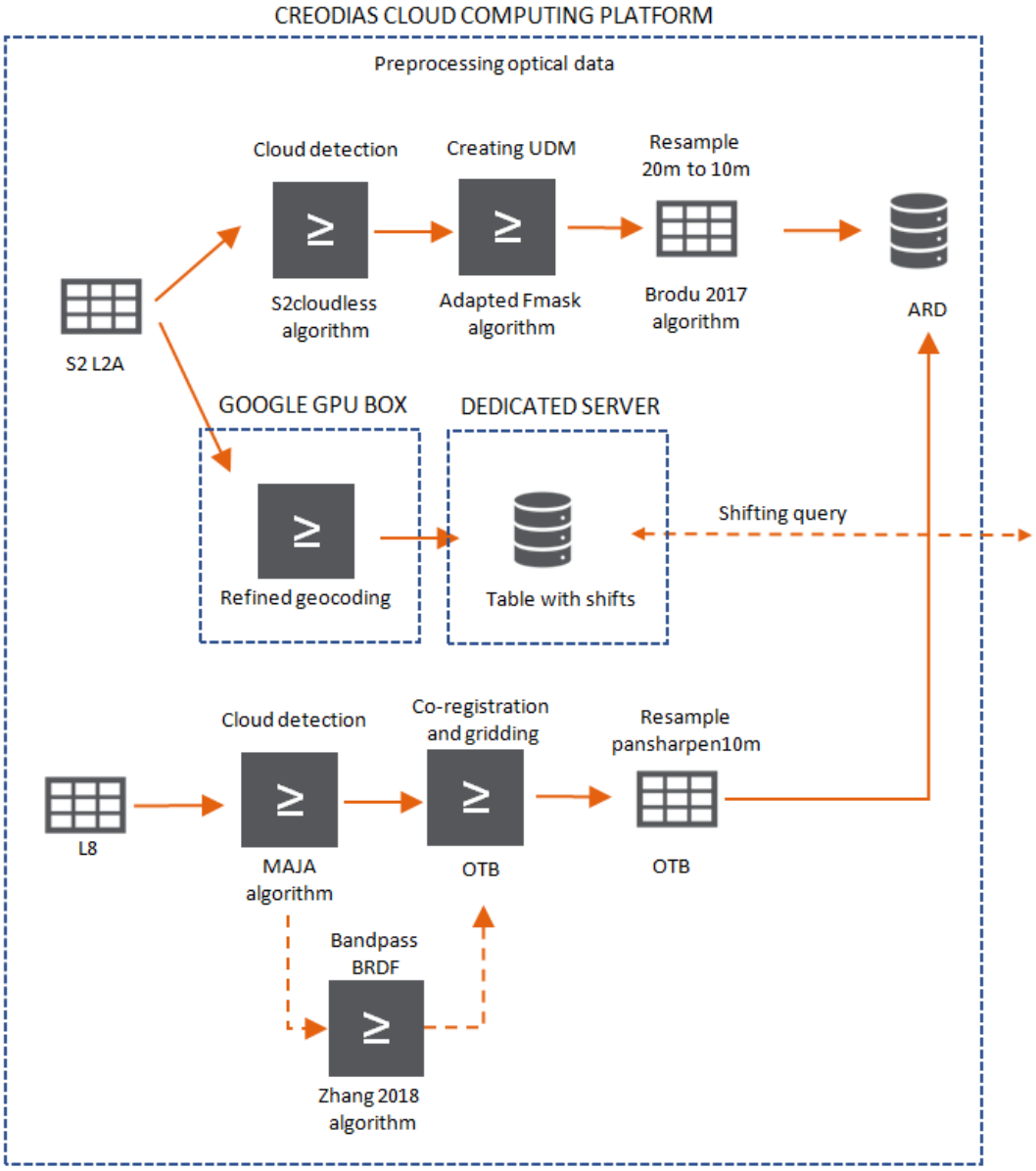


Figure 19: Overview of Sentinel-2 (and Landsat-8) pre-processing

For an overview and mapping total cropped area, cloud-free composites can be an adequate input. Based on ARD, we will provide monthly cloud-free composite based on the Weighted Average Synthesis Processor (WASP) algorithm (Hagolle et al., 2018) developed during the ESA Sentinel-2 for Agriculture project and adapted to S2-ARD and Landsat-8 ARD data. These cloud-free composite will be compared to the data provided by the Copernicus Sentinel-2 Global Mosaic (S2GM) service[11] during a small benchmark phase.

To sum up, the optical High Resolution (HR) ARD data will have the following characteristics:

- Pixel content: Surface reflectance

[11] https://land.copernicus.eu/imagery-in-situ/global-image-mosaics/node/16

- Spatial resolution: 10 metres
- Geographic projection / reference system: UTM 37S (EPSG: 32737)
- Gridding: MGRS grid
- Spectral characteristics: 10 bands for Sentinel-2 and 6 bands for Landsat 8
- UDM mask with cloud, snow, pixel quality
- GeoTIFF format for raster and mask

This dataset will be accompanied by:

- A set of monthly cloud-free composite over the vegetation period
- A set of functions to compute on-the-fly the main radiometric indices from the ARD products

Additional high resolution optical data can be integrated in this dataset for specific areas where the cloud coverage can decrease the density of the time series. For example, it may be possible to get access to Planet data that is made freely available in monthly cloud-free composites through the Norway's International Climate & Forests Initiative (NICFI). Planet data does not record the Short-Wave Infra-Red (SWIR) band that is available on Sentinel-2, but the high resolution (3.7m) can provide additional information.

For the optical scenes with small clusters of missing values due to clouds, values can be restored with SOMs as developed by Kohonen et al. (2001). When filling clusters with missing data is not feasible or useful with SOMs due to large gaps, values at a certain time-interval are derived from previous and later reads. These steps can be done to ensure a continuous temporal signal with a fixed interval.

Annex V.High Spatial Resolution SAR data: Sentinel 1

Sentinel-1 mission is composed of two identical SAR satellites (1A and 1B) operating at C-band (5.405 GHz, 5.6 cm wavelength). Due to the 12-day revisit, this mission offers the possibility to have a 6-day revisit maximum in full-operation scenario (when both satellites cover the same area). The acquisition scenario is driven by policy and science requirements. For land surface, the satellites operate in the Interferometric Wide swath (IW) mode with dual polarisation acquisition. The density of IW acquisitions is optimised over European area. Sentinel-1 time series will be used to discriminate land cover classes further based on their temporal behaviour (backscatter), consequently improving accuracy in areas with an extensive amount of cloud cover. Copernicus generally provides level-1 products Single-Look Complex (SLC) and Ground Range Detected (GRD). In many cases, level-0 (Annotated Raw) data is available for users who want to pre-process SAR by themselves.

Sentinel-1 image processing, from level 1.1 SLC product to a calibrated map product, will be done on virtual machines running ESA SNAP (Sentinel Application Platform), using tested SNAP processing graphs to run the required processing steps for each processing chain. The outputs will be the calibrated Sentinel-1 VV and VH backscatter images, polarimetric entropy and alpha, geometry information (local and global incidence angles) and meta-information (map projection, grid size, etc.). Interferometric coherence requires two SLC input images with the same incidence angle and path. To minimise temporal decorrelation, acquisition time interval should be minimum, preferably 6 days when both Sentinel-1A and Sentinel-1B are available. This only occurs in some areas North of the Equator; most areas are recorded with 12-day interval. Interferometric processing needs precise sub-pixel co-registration and phase preservation of the two input images. Combined Sentinel-1A and -1B interferometry, only for the 6-day areas, requires along-track alignment / mosaicking (called Slice Assembly in SNAP) in order to ensure sufficient overlap.

We propose three SLC image processing procedures for the different outputs: a) the VV, VH backscatter map products, b) dual-pol polarimetric output and c) the interferometric coherence calculation. The third one requires two SLC images as input with (almost) the same overlap.

a) Backscatter ortho product: calibrated, ortho- and map-projected backscatter. SLC processing requires some extra handling of sub-swaths in respect to GRD input, but ortho-output will be the same (see Figure 20, left).

b) Polarimetric processing. Sentinel-1 is not a full polarimetric sensor, as it has only two polarisations. Nevertheless, VV and VH phases are related, making dual-polarisation polarimetry possible. The SNAP processing provides polarimetric tools for calculating a 2 x 2 scattering matrix (C2) and dual-pol Entropy(H)-Anisotropy-Alpha decomposition. These parameters give information about the dominant scattering mechanism and type of scatter. This information improves image classification (see Figure 20, right).

c) Interferometric processing calculates phase differences between two overlapping SLC images from the same satellite path. Interferometric coherence is an indicator for crop growth or cultivation changes (see Figure 21).

S1 SLC ortho
· Read image and metadata · Apply precise orbit · Radiometric Calibration · Terrain Observation with Progressive Scans (TOPS) Deburst · Multi-looking · Terrain Correction

S1 Dual-pol polarimetry
· Read image and metadata · Apply precise orbit · Radiometric Calibration · TOPS Deburst · Polarimetric C2 Matrix Generation · Polarimetric Speckle Filter · Polarimetric H-A-Alpha decomposition · Multi-looking · Terrain Correction

Figure 20: Sentinel-1 IW single image processing flows: SLC calibration and ortho-processing chain (left) and calculation of polarimetric entropy, alpha and anisotropy (right).

S1 SLC bi-temporal coherence	
Image [t]	**Image [t-1]**
· Read image and metadata · Apply precise orbit · TOPS Split · [Slice Assembly]	· Read image and metadata · Apply precise orbit · TOPS Split
· Coregistration · Coherence calculation · TOPS Deburst	
· TOPS Merge · [Multi-looking · Terrain Correction	

Figure 21: Sentinel-1 IW interferometric coherence calculation: splitting input images into sub swaths, apply Slice Assembly mosaicking if necessary (for full overlap), co-register sub swath pairs, calculate inSAR coherence, merge sub swath coherence images and apply ortho-processing

All products from the processing flows acquired during the growing seasons provide signatures that give temporal information. Overall statistics, such as temporal minimum, maximum, mean (see Figure 22) and standard deviation of backscatter (see Figure 23), polarimetric parameters and coherence can help discriminating various land cover types, cropping dynamics, inundation/flooding information.

Figure 22: Example of multi-temporal min-mean-max of all VH images acquired during the long wet season March-August 2019 over the Nairobi area

Figure 23: Example of multi-temporal standard deviations of VV and VH images acquired during the long-wet season March-August 2019 and the small wet season Oct-Dec2019 over the Nairobi area

The variations of radar (and optical) parameters may give useful information of particular phases during the crop growing season: soil preparation, increase of biomass, flowering, ripening and harvest. As an example, optical (Normalized Difference Vegetation Index (NDVI)) and radar (VH/VV, VH and VV) temporal signatures of some crops are shown in Figure 24. These crops show a clear start and end of their growth. Unfortunately, not every crop shows such clear temporal behaviour, this depends on crop type and how open the vegetation structure is. Open structures with a large contribution of the ground component show a less explicit signature.

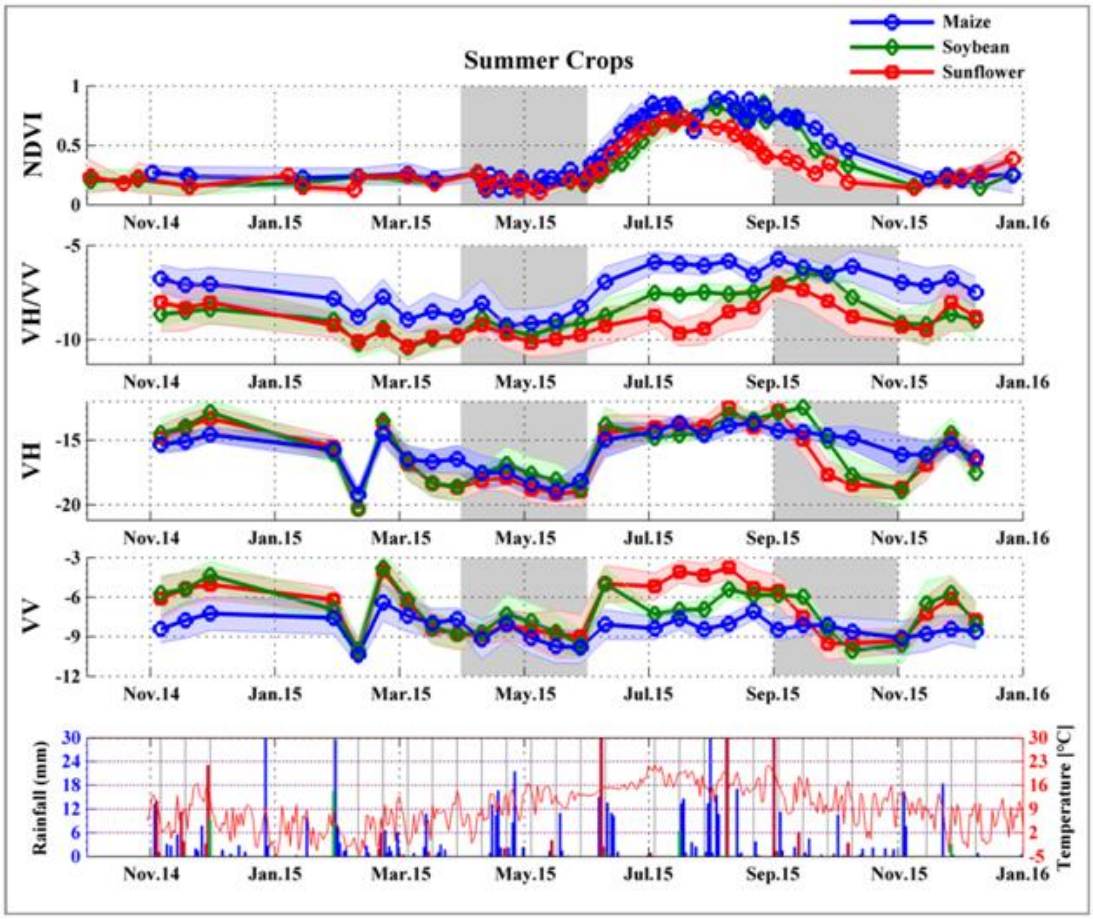


Figure 24: NDVI and SAR signatures of some crops in France. NDVI, VV and VH show a clear increase in June and a decrease in October, related to crop growth. Other backscatter changes are not related to biomass but to soil roughness, wetness and land cultivation (Feb-Mar) (Veloso et al., 2017).

Annex VI. Crop Type Classification method

The different approaches available for the classification of crop types from available satellite imagery are described in the following paragraphs.

Basic supervised classification

This method can be applied in low complex situations with only a few main crops to be identified and where crop patterns are simple *e.g.,* large areas of a single crop. The first step in a supervised classification is the identification of training data for the required classes from reference data collected on the field. Based on the training data, spectral signatures are calculated for each class. These class spectral signatures are then used to classify the complete image or scene. The most common supervised classification methodology is Maximum Likelihood (ML). ML is widely used and is a parametric classifier that is based on statistical theory. More advanced classifiers such as RF (Random Forest) and SVM (Support Vector Machine) are available as well, which are both machine learning algorithms. RF is the preferred method as high accuracy is generally reached with this algorithm as well as genericity. Many crops have a unique spectral (optical) or backscatter (SAR) temporal signature that allows for discrimination of various crops within an AOI. Rice and corn are, for example crops which are easy to detect with a degree of high accuracy with this method. In this case, a "traditional" pixel-based supervised classification is sufficient and saves precious time, because no effort is needed to create additional training data for machine learning algorithms. If the area is more complex (e.g., many crops, smaller fields) machine learning algorithms like RF will be applied.

Advanced supervised classification: Machine learning

In complex cropping environments (e.g. many crop types, small fields, inter-cropping), more advanced classification methods can yield better results than basic supervised classification methods. There are various machine learning (ML) algorithms available to the consortium that can be applied operationally. We have a Python-based Temp-CCN method as well as various ML methods in ORFEO Toolbox called IOTA2.

The temporal CNN (originally developed by Pelletier et al. (2019b)) can be deployed with both pixel-based and object-based inputs. This model uses the temporal signal of pixels in sampled crop fields in a growing season as training data and applies a Deep Learning (DL) architecture including several convolutional layers (see Figure 25).

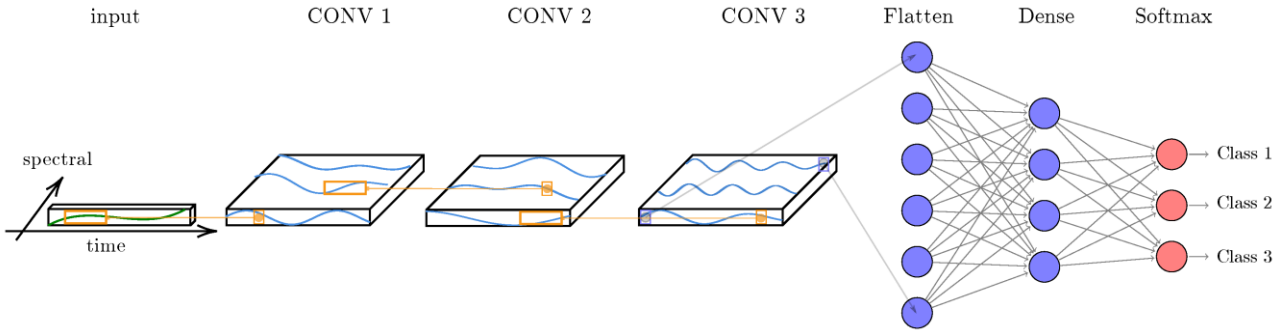


Figure 25: Temporal Convolutional Neural Network (TempCNN) proposed by (Pelletier et al., 2019b) mixing spectral and temporal convolution to improve crop mapping

The input for the DL architecture will be generated by the fieldwork efforts and applied on a stack of unlabelled satellite imagery to predict the crop types. TerraSphere has implemented this method to map complex cropping patterns in parts of Myanmar in 2018 and 2019. Another example of a large-scale

implementation of this method is the Monash Vegetation Map where the methodology was successfully applied for an area of 227,000 km² in Australia (Pelletier et al., 2019a).

IOTA2 is a processing chain which uses the Orfeo Tool Box main steps for its classification process. It allows a choice of classifiers like Support Vector Machine (SVM) and Randon Forest (RF). A ratio can be chosen that will split the fieldwork dataset into training & validation. IOTA2 calculates many different indices (e.g. NDVI, NBUI, NDWI) to add to the time-series of Sentinel-2 data. A number of other global parameters can be set, like gap filling (allowing to fill the no-data areas on your S2 acquisitions with other S2 acquisitions, to obtain continuous time series), and auto context (which adds a segmentation step to the process).

All available ML algorithms as described above are thoroughly tested on parts of the Kenya AOI during the 2021-2022 growing season, and the IOTA2 (RF) classifier continuously yielded the best results. Therefore, IOTA2 is currently our default classifier for the coming 2022-2023 growing seasons in Kenya.

Object-based classification

Segmentation of satellite data into objects can be a first step for both previous methodologies. It can be applied best in case enough optical data is available. This is explained by the fact that segmentation results on SAR data do not generally yield good results in cases of small field sizes (as can be expected in the AOI's) due to speckle phenomenon. It also works best in less complex cropping patterns where there are large areas of single crops. Object-based classification can enhance crop-mapping since signals within the object can be grouped and outliers within objects are filtered. This leads in general to smoother and better crop maps with less "noise" of misclassified individual pixels. The first step is to segment pixels into vectorized boundaries representing a crop field, or a group of fields with the same crop. These polygons can then be used to get aggregated (field) data and feed this data to classification algorithms. In order to train a powerful model, much sample data is needed in terms of labelled polygons. Examples of open-source implementations of such segmentation can be found here[12]. The results of these different steps will be combined in one map. A more technical description is given in Persello et al. (2019).

Considering the expected field size that are expected to be rather small as compared to the Sentinel-2 spatial resolution, we do not anticipate applying an object-based approach over the entire AOI. However, a segmentation model will be applied to facilitate the visual interpretation of the segments prior to the field work. The segmentation model can be trained based on our in-house Mask-RCNN, based on original Mask-RCNN (He et al., 2017) (see Figure 26), which runs on Sentinel-2 chipped imagery. In the case of insufficient data availability, an already trained segmentation algorithm can be run and validated where it yields results in the contexts of the AOIs.

[12] https://www.analyticsvidhya.com/blog/2019/07/computer-vision-implementing-mask-r-cnn-image-segmentation/

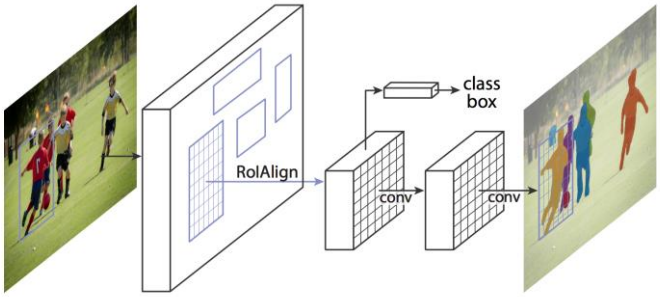


Figure 26: Mask R-CNN framework for instance segmentation (He et al., 2017)

Mitigation measures for prolonged cloudiness

Limited availability of optical satellite data (Sentinel-2/Landsat-8) can be expected for some AOIs. Yet we aim to use every bit of cloud-free image part because the MMU of 0.04 ha is largely dependent on it. The use of Self-Organising Maps (SOMs) helps to make cloud-free composites by filling gaps with interpolated values. The use of SAR (Sentinel-1) data is another method that allows for crop mapping when little optical data is available.

Self-Organising Maps

Missing optical scene parts due to clouds and shadows can be restored with SOMs. This can be an effective strategy when there is enough optical data available in the vicinity (temporally) of the image to be restored, or if other data sources (Landsat-8) are available. This will allow for a temporal denser time series of cloud-free optical data as a basis for classification. The consortium have applied this method successfully during the previous growing seasons in Kenya, where nearly 100% cloud-free mosaics were produced for each growing season. This even made the use of Sentinel-1 data redundant for those seasons. As each season is unique, we realise the situation with respect to cloudiness can be less favourable for the next seasons.

The use of SAR data

Sentinel-1 data are always available throughout the entire growing season. Besides the guaranteed availability, some studies even indicate that a multi-sensor approach (combining S1 and S2) yields better results in crop mapping than using only S1 or S2 (Kussul et al., 2015; Van Tricht et al., 2018). The Sentinel-1 derived data is two-fold: (1) continuous and smoothed Gamma0 backscatter signals to generate a temporal signature in VV and VH and (2) temporal coherence markers indicating change. The change can be due to management like ploughing, harvesting or due to crop phenological characteristics like crop emergence, flowering or senescence. These extra signals help to better discriminate between crops. Therefore, we anticipate the use of SAR data not only in cases where there is no alternative, but also as additional data source. The combined use of S1 and S2 is possible in the various classification methodologies described above, where S1 backscatter as well as coherence can be additional layers in the data stack to be classified.

Annex VII. Validation strategy

The last step in the production of the crop mapping consists in analysing the collected field samples in order to draw conclusions about the accuracy of the products (in-season and end-of-season crop masks and crop type maps). Since the sampling presented in section 2 is well spatially distributed over the study area, segments can randomly split into training data for the classification (75%) and validation (25%) sets. This split is done at the field parcel level (i.e., within segments) and not on individual pixels, to ensure the independence of the validation samples. The proportion of validation samples may be increased if sufficient training data is obtained.

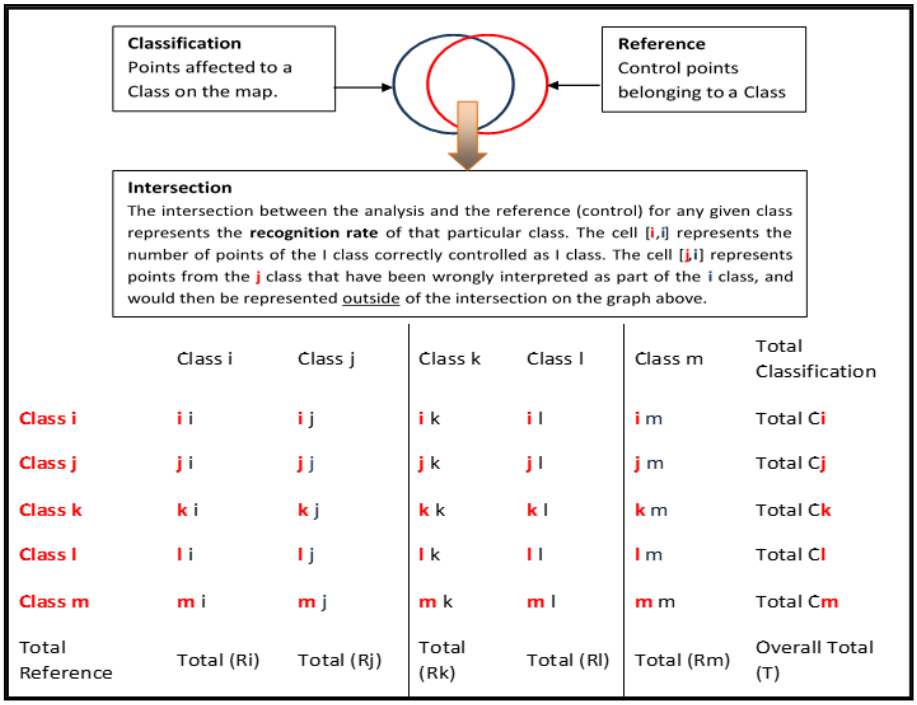
Thematic accuracy is presented in the form of an error matrix made from the results of the sample units visited in the field or visually interpreted for non-cropped areas. As explained in Selkowitz and Stehman (2011), the sampling intensity (defined as the proportion of the population that is been sampled, not to be confused with sampling size that refers to the number of sample units) resulting from the stratified systematic sampling approach is not equal for all strata. These differences should be accounted for by applying a weight factor to each sample unit within a given stratum. Therefore, a correction for the sampling intensity will be applied to the error matrices produced following the procedure described by Selkowitz and Stehman (2011) and applied by Olofsson et al. (2013) leading to a weighting factor inversely proportional to the inclusion probability of samples from a given stratum. Not applying this correction could result in underestimating or overestimating map accuracies. This was already applied successfully as part of the Copernicus Land lot 1 External Validation Contract.

In addition, sampling weight factors are also needed to correct for imperfections in the sampling scheme that might lead to bias between the number of samples and the size of the stratum (reference population). In other words, the purposes of the weight factors are to compensate for unequal probabilities of selection and to compensate initial sampled units that could not be surveyed due to accessibility issues. For example, this correction is required because there are less samples in the “other areas” strata (i.e., lower sampling intensity) than those in the “cropland” strata for an equivalent stratum area. Once the field campaign is finalised, final sampling intensities will be calculated for each stratum based on the ratio between the number of samples included in the survey and the size/total area of the stratum considered and the weight factor (p) will be calculated according to the following equation:

$$\hat{p}_{ij} = \left(\frac{1}{N}\right) \sum_{x \in (i,j)} \frac{1}{\pi_h^*}$$

Where i and j are the columns and rows in the error matrix, N is the total number of possible units (population) and π_h^* is the sampling intensity for a given stratum.

Thematic accuracy is usually assessed based on the construction of a confusion or error matrix for which it is crucial that unequal sampling intensities as described above are accounted for as described by Olofsson et al. (2014). Then, the construction of the error matrix is described as illustrated in Figure 12 for 5 thematic classes.



	Class i	Class j	Class k	Class l	Class m	Total Classification
Class i	i i	i j	i k	i l	i m	Total Ci
Class j	j i	j j	j k	j l	j m	Total Cj
Class k	k i	k j	k k	k l	k m	Total Ck
Class l	l i	l j	l k	l l	l m	Total Cl
Class m	m i	m j	m k	m l	m m	Total Cm
Total Reference	Total (Ri)	Total (Rj)	Total (Rk)	Total (Rl)	Total (Rm)	Overall Total (T)

Figure 27: Confusion Matrix for Accuracy Assessment of thematic map product

Referring toFigure 27, let us assume α represents any given class of [*i, j, k, l, m*], the following accuracy metrics can be calculated:

- The **Overall Accuracy** or **Recognition Rate** is measured by the sum of the diagonal of the Confusion Matrix divided by the total number of controlled points: $OA\ or\ R_r = \sum_{\alpha=i}^{m}(\alpha\alpha)/\mathrm{T}$. The **Recognition Rate** or **Overall Accuracy** assesses the overall agreement between the classified and reference data set. However, for single class themes, it does not necessarily provide a realistic assessment of the quality of the map produced because there can be substantial unbalance between omission and commission errors.
- Therefore, the row and column totals and the diagonal of the Matrix are used to assess two types of accuracy, the User's and Producer's Accuracy:
    - **Producer Accuracy** for the α class = $\alpha\alpha/\mathrm{C}\alpha$ is a measure of **omission error**. For instance, an observation has been identified as Tree Covered in the validation set, but has been classified as another class: it has been omitted from the forest class..
    - **User Accuracy** for the α class = $\alpha\alpha/R\alpha$ is a measure of the **commission error** (or contamination risk): errors due to the wrong allocation of an observation to a class. For instance, an observation is classified as forest, but identified as belonging to another class during the validation process: this observation has contaminated another class.

The standard error of the overall, producer and user accuracies can be calculated as follows:

$$\sigma_h = \sqrt{Var_h} = \sqrt{\frac{p_h(1-p_h)}{n_h}}$$

where n_h is the sample size for stratum h and p_h is the expected error rate.

The total standard error is then the square root of the sum of the variance times the square of the area D of each stratum:

$$\sigma_{Total} = \sqrt{D_h^2 . Var_h}$$

The 95% confidence interval ($CI_{95\%}$) which indicates that the "true " value has a 95% probability to be included within the calculated range, is then defined as:

$$CI_{95\%} = \pm 1.96 * \sigma_{Total}$$

The same approach will be adopted for the in-season as well as end-of-season crop masks and crop types maps.

Annex VIII. Area estimate bias correction

Crop area estimates can be derived directly from the field data alone using the so-called direct expansion method as long as the data has been collected based on a probabilistic sample or for which a suitable method was used to correct any potential bias. Therefore, early area estimates (Direct expansion estimators) can be provided as soon as the results from the field campaign have been collated and analysed event before the classification of the satellite imagery.

Assuming a probabilistic sample or a sample for which the bias has been corrected following the formulae described in the previous annex in which a weight factor is applied to each sample unit to ensure sampling intensity has been accounted for, the estimate of proportion (y) of class (c) and its variance are given by:

$$\bar{y}_c = \sum_{i=1}^{n} \frac{1}{n} y_i \text{ and } var(\bar{y}_c) = \left(1 - \frac{n}{N}\right) \frac{1}{n(n-1)} \sum_{i=1}^{n} (y_i - \bar{y}_c)^2$$

where: y_i is the proportion of segment *i* covered by class *c*, *N* is total number of segments in the region, *n* is number of segments in the sample. The proportion of the study region sampled (*n/N)* is the sample fraction. The estimate of class area (Z) and variance in study area (D) are as follows:

$$\hat{Z}_c = D * \bar{y}_c \text{ and } var(\hat{Z}_c) = D^2 * var(\bar{y}_c)$$

where D is the area of the stratum. It is better to compute the estimates first as proportions rather than as absolute areas because this automatically takes account of errors resulting from small, localised variations in the scale of segment maps and drawing or digitising errors. The Direct expansion estimators are defined based on the results of the field campaign alone and calculated for each stratum present in the AOI. The total estimate just corresponds to the weighted average of the proportions according to the area covered by each stratum. The standard error for the whole area is then the square root of the sum of the variance times the square of the area for each stratum:

$$\sigma_{Total} = \sqrt{D_h^2 . Var_h}$$

where D_h is the stratum area. The 95% confidence interval is +/- 1.96. σ_{Total}.

However, the confidence interval of the estimate derived from direct expansion is likely to be relatively large. To improve the precision of the estimates, field segment data can be combined with classified satellite imagery. In this latter case (i.e., using the classification map), a so-called Model Assisted Regression (MAR) estimator can be applied. MAR is more reliable than any other area estimation methodology as it provides both an area estimation per cover type together with an indication of its uncertainty.

In brief, MAR relies on the combination of area estimates made at the segment level for both ground data and classified satellite imagery. The observation is paired, and a regression analysis is performed as illustrated in Figure 28.

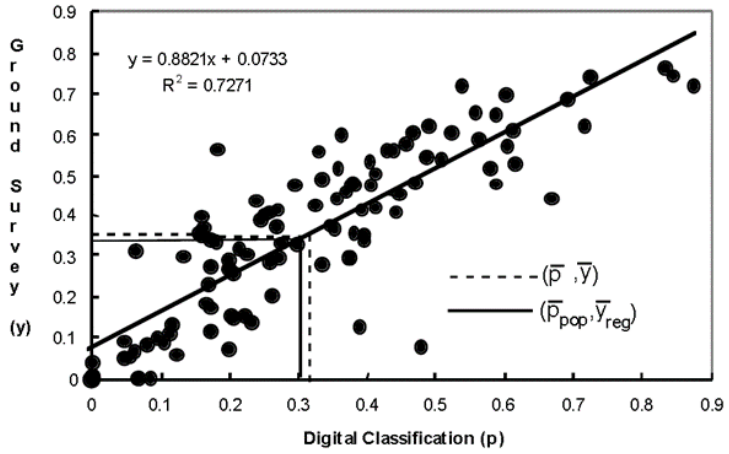


Figure 28: Relationship between the proportion of wheat within each segment from the digital classification (p) and ground survey (y) in the UK (after Taylor, 1997).

The MAR estimator methodology is fully described in Taylor (1997). The MAR estimator y_{reg} is calculated based on the following equation:

$$y_{reg} = \bar{y} + b * (\bar{p}_{pop} - \bar{p})$$

where $\bar{y}$ is the mean field data sample value, b is the slope of the regression line, $\bar{p}_{pop}$ is the proportion of pixels classified as the crop in the whole of the region of interest and $\bar{p}$ is the classified image mean sample value. The variance of the estimate is calculated as:

$$var(\bar{y}_{reg}) = \frac{1}{n} var(y)(1 - r_{py}^2)$$

Where r_{py}^2 is the regression coefficient.

Therefore, the higher the regression coefficient the smaller the variance and as a result the precision of the estimate. Experience for the MARS programme (Taylor, 1997) showed that high classification accuracy was correlated with high regression coefficient.

The estimation of land cover type areas obtained by such procedure can be very variable from pixel counts because image classification is affected by misclassification errors affecting the classes. Area estimates derived from the MAR estimator method are corrected from misclassification errors whilst exhibiting a more precise estimate than that of the direct expansion estimate thanks to the complete coverage provided by the image classification.

In summary, direct expansion estimates are unbiased, but suffer from high sampling error, pixel counts from classified satellite imagery are biased but have no sampling errors and the combination of ground data and classified imagery are unbiased and exhibit a reduced sampling error.

The efficiency of the MAR estimator is estimated by the relative efficiency (n_{reg}), which is the ratio of the variance from the MAR estimator method and the direct expansion estimate which simplifies to:

$$n_{reg} = \frac{1}{1 - r_{py}^2}$$

During the MARS project, it was shown that relative efficiency above 2 could be obtained with single date imagery. Nowadays, thanks to the availability of multi-temporal image coverage, higher classification accuracy could be achieved resulting in reduction of variance by a greater factor. For land cover types that

are very distinct such as tropical rainforest, very high relative efficiency (i.e. reduction of the uncertainty) can be obtained with the Regression Estimator as shown by Sannier et al. (2016). In their study of forest cover in Gabon the Authors reported relative efficiency close to 60 as illustrated in Figure 29 in which there is a substantial reduction in the error bars shown when comparing the direct expansion with the MAR estimates. In other words, to obtain the same level of uncertainty with the direct expansion *i.e.* without using the EO-based forest cover map, the sample size would need to be increased by a factor of 60. The 95% confidence interval of the direct expansion estimate is around 2% of the total area of Gabon when it is reduced to 0.25% for the regression estimate. It is also worth noting that this is a rare case for which the pixel count from the map is relatively close to unbiased estimates considering that it is invariably contained within the 95% confidence intervals bounds of the, most precise, MAR estimates.

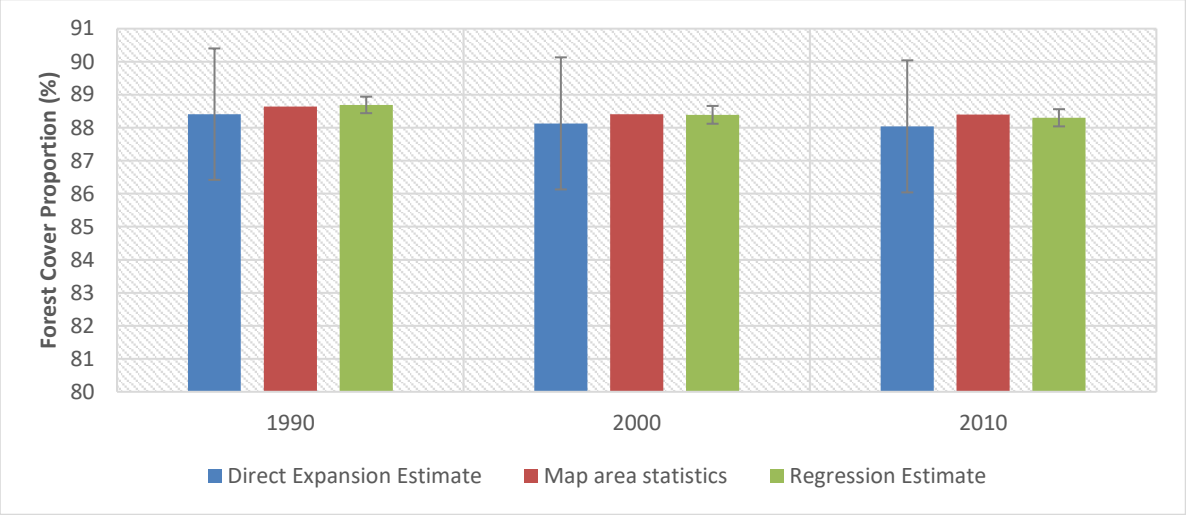
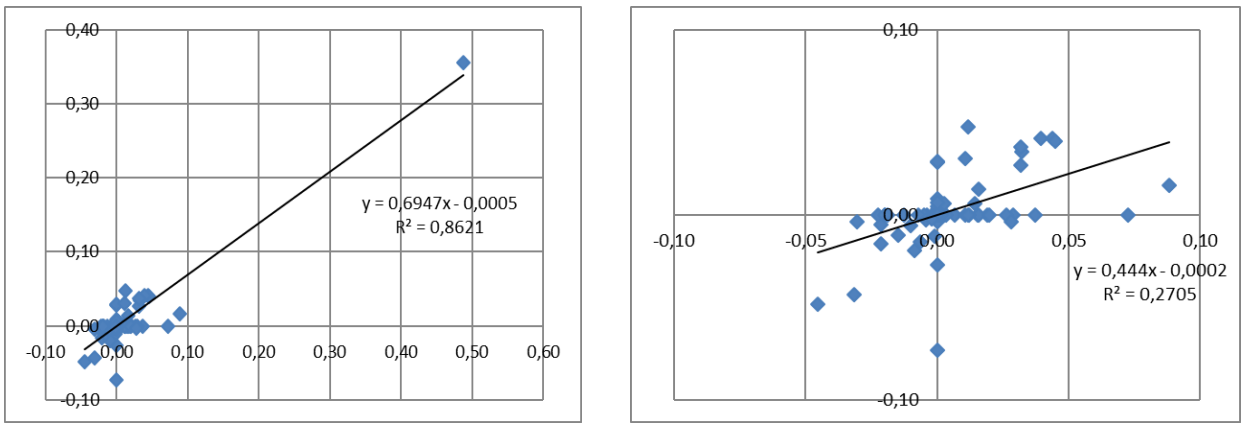


Figure 29: Comparison of Direct Expansion, Map area Statistics and Regression estimates for forest cover in Gabon in 1990, 2000 and 2010 and associated 95% confidence intervals as error bars.

One of the issues with the regression estimator as described above is that it is potentially very sensitive to the quality of the linear regression which is sometimes poorly reflected by the R^2 as shown in the example below (see Figure 30).



(a) (b)

Figure 30: Illustration of the importance of the quality of the linear regression with (a) all observations and (b) one observation removed (X and Y axes are respectively the proportion of the land cover for each segment in the ground and classified data).

A slightly adapted estimator is also available in the form of the Generalised Regression (GREG) estimator described by (Särndal et al., 1992, sec. 6.5) which focuses on the average of the differences from each observation to the mean of observations:

$$\hat{\mu}_{greg} = \frac{1}{N}\sum_{i=1}^{n}\hat{y}_i - \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)$$

with variance:

$$\widehat{var}(\hat{\mu}_{greg}) = \frac{1}{n(n-1)}\sum_{i=1}^{n}(\varepsilon_i - \bar{\varepsilon})^2$$

where N is the number of map units, n is the reference set sample size, y_i is the observation for the ith reference set sample unit, $\hat{y}_i$ is the map class, $\varepsilon = \hat{y}_i - y_i$, and $\bar{\varepsilon} = \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i$.

This latter estimator has the advantage of not being sensitive to the quality of the linear regression but could potentially provide an estimate with a larger confidence interval compared with the direct method when the EO-based map is of poor quality for the crop type considered. Such an approach was applied by Sannier et al. (2014) for estimating forest area and forest area change but could also be easily applied to crop types.

Therefore, considering all the factors outlined previously in the previous sections, the following approach will be adopted for providing crop area estimates depending on the level of complexity as described in the feasibility study (D1.1) :

- Direct expansion and Regression estimates as outlined by Taylor (1997) will be applied based on the surveyed segments
- In case a substantial number of segments cannot be surveyed, a bias correction adapted from Brand et al. (1999) and Simms et al. (2016) may be applied, but it should still be possible to provide a direct expansion and Regression estimates as outlined by Taylor (1997) once a bias correction has been applied

Finally, it should also be stressed, that three sets of crop area estimates will be provided instead of two as laid out in the tender specifications:

- Direct expansion estimates as soon as the field campaign has been completed.
- In-season improved crop area estimates integrating the in-season maps.
- End-of-season final improved crop area estimates integrating the end-of-season map.