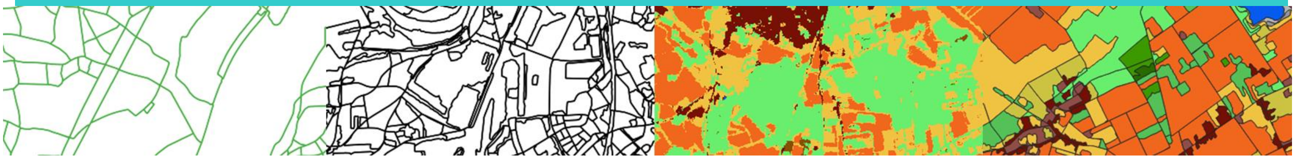
Service contract for the Copernicus Land monitoring services

Crop Mapping for GEOGLAM Country Level Support



Framework contract 939708-2020-IPR

In-season Crop Type Map & Crop Mask

Kenya - long rains season – 2023

Prepared by:

CLS

COLLECTE LOCALISATION SATELLITES

&

TerraSphere

Reference : In-season mapping - Kenya - long rains season 2023

Issue 1.0 - 22/09/2023

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	22-09-2023	First version submitted to JRC

1 Introduction

This document describes the in-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 and the ground truth data used for the classification. The document only describes in detail the fieldwork and satellite data pre-, and post-processing as far as they are different from what has been described in detail in the feasibility study for Kenya.

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) May-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 data used

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

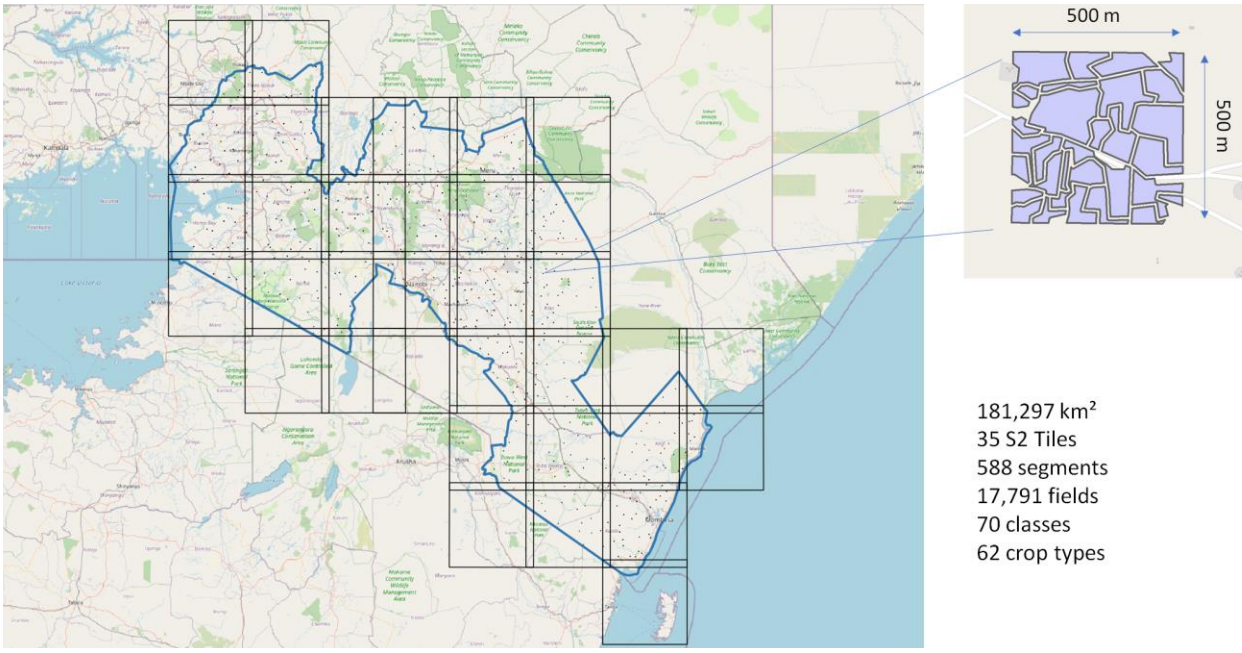


Figure 2. Kenya AOI overlaid with the S2 tile-based grid and the fieldwork segments

2.1 Satellite data – Sentinel-2

In total, 1,132 Sentinel-2A & B Level-2A images have been acquired covering 35 tiles between 01-04-2023 and 15-07-2023. 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	01/04/2022	15/07/2023	21
36MXE	01/04/2022	15/07/2023	21
36MYC	01/04/2022	15/07/2023	42
36MYD	01/04/2022	15/07/2023	21
36MYE	01/04/2022	15/07/2023	21
36MZC	01/04/2022	15/07/2023	42
36MZD	01/04/2022	15/07/2023	42
36MZE	01/04/2022	15/07/2023	42
36NXF	01/04/2022	15/07/2023	41
36NXG	01/04/2022	15/07/2023	41

Tile ID	First Date	Last Date	Number of Images
36NYF	01/04/2022	15/07/2023	21
36NYG	01/04/2022	15/07/2023	21
36NZF	01/04/2022	15/07/2023	42
37MBT	03/04/2022	12/07/2023	21
37MBU	03/04/2022	12/07/2023	21
37MBV	01/04/2022	15/07/2023	42
37MCR	03/04/2022	14/07/2023	42
37MCS	03/04/2022	14/07/2023	42
37MCT	03/04/2022	14/07/2023	42
37MCU	03/04/2022	12/07/2023	21
37MCV	03/04/2022	12/07/2023	21
37MDR	05/04/2022	14/07/2023	21
37MDS	03/04/2022	14/07/2023	42
37MDT	03/04/2022	14/07/2023	42
37MDU	03/04/2022	14/07/2023	42
37MDV	03/04/2022	12/07/2023	42
37MEQ	05/04/2022	14/07/2023	21
37MER	05/04/2022	14/07/2023	21
37MES	05/04/2022	09/07/2023	21
37MET	05/04/2022	14/07/2023	21
37MFS	05/04/2022	14/07/2023	42
37MFT	05/04/2022	09/07/2023	42
37NBA	01/04/2022	15/07/2023	42
37NCA	03/04/2022	12/07/2023	21
37NDA	03/04/2022	12/07/2023	42

2.2 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%) & 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 digitised segments and landcover classes (amongst others) are recorded. It should be noticed that some segments have not been visited in the field due to the absence of crops or for the safety of the enumerators. To create an input for classification, point data are assigned to the polygons. In the case of no point is recorded (due to e.g. inaccessibility of segment), the land cover class recorded during the first digitising of the segments prior to the field campaign, is automatically assigned. The polygons labelled “cropland” not surveyed (initially supposed to be) 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.

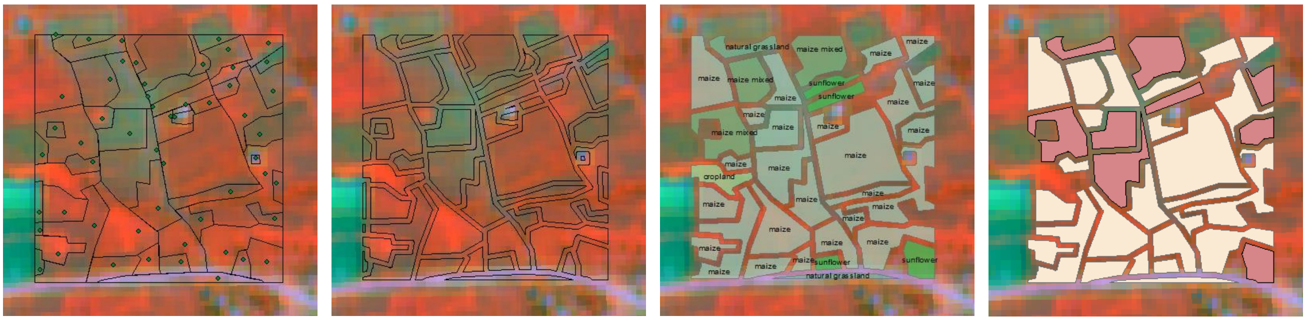
2) A negative buffer of 5 meters is applied to eliminate, or at least minimize, the boundary effects between different classes that will negatively impact the purity of training samples signatures. Consequently, polygons are always separated by 10 meters, which corresponds to the size of 1 Sentinel-2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

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. Indeed, as explained is section 0, the classification workflow is applied per S2-based block.

The Figure 3 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 3. 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 63 individual crop types. The figures below show a few examples from the fieldwork campaign.

Figure 4. Field with maize in monoculture

Figure 5. Field with maize and cabbage in mixed-cropping

Deviations from feasibility study proposal and the short rains season:

There were no deviations from what was described in the feasibility study and what was done during the short rains season 2022-2023.

3 Workflow

3.1 Pre-processing

The pre-processing of the satellite data applied was unchanged from what was proposed in the feasibility study (D1.1). For each of the two satellite data types some specific pre-processing are summarised as follows below.

Sentinel-2

Based on the Sentinel-2 L2A data, we reprocessed the cloud masks using S2cloudless and Fmask algorithms for detailed removal of clouds and cloud shadows. Monthly syntheses are then processed using the WASP algorithm (open-source solution developed by CNES[2]). 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 6 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.

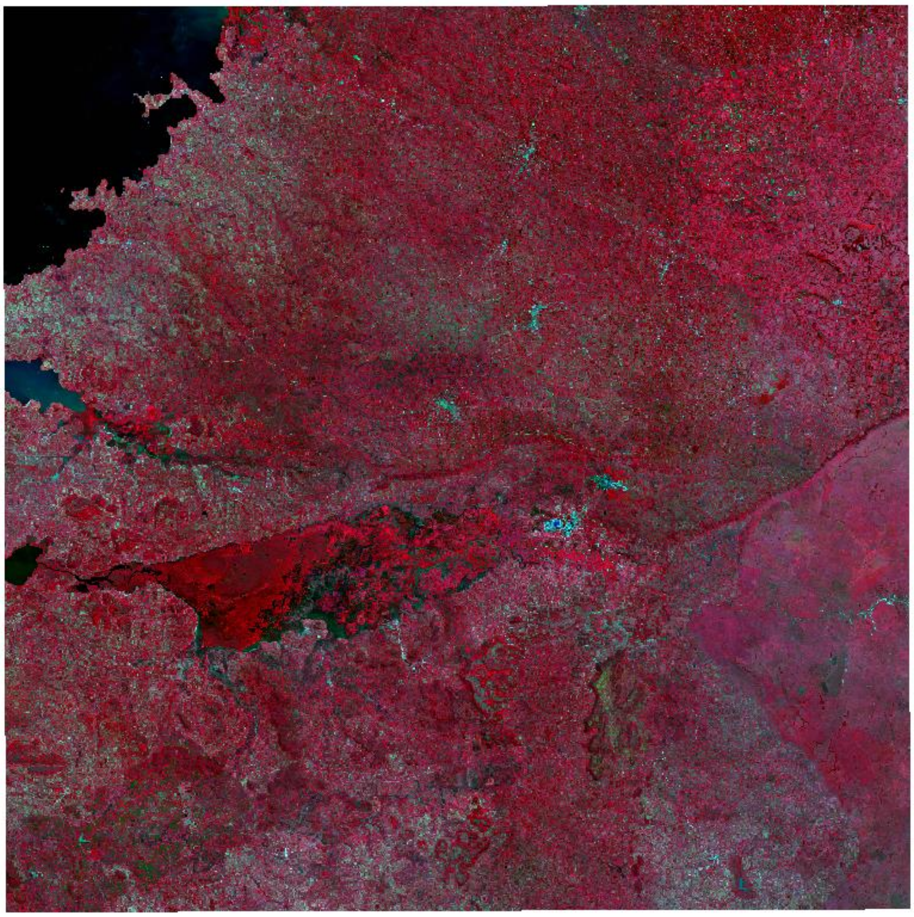


Figure 6: Sentinel-2 monthly synthesis composite, 15/05/2023, tile 36MXD.

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

Based on these monthly synthesis, four spectral indices are computed: the Weighted Difference Vegetation Index (WDVI[3]), Normalized Difference Vegetation Index (NDVI), Normalized Difference Water Index (NDWI), and Brightness Index (BI). All layers are used as input in the classification algorithm.

Landsat-8

The use of the Landsat-8 dataset was not considered as 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.

Deviations from feasibility study proposal and the short rains season:

There were no deviations from what was described in the feasibility study and what was done during the short rains season 2022-2023.

3.2 Classification

Crop Type – Various classification algorithms were tested during the previous services, including supervised (maximum likelihood) classification, TempCCN and Random Forest (RF). It was decided to use the RF classification as final method for the Kenyan long rains season mapping 2023. The algorithm is characterized by relatively simple parameterization, a good computation efficiency, and highest accuracy. 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 Figure 7 shows the result of the raw classification output, before post-processing.

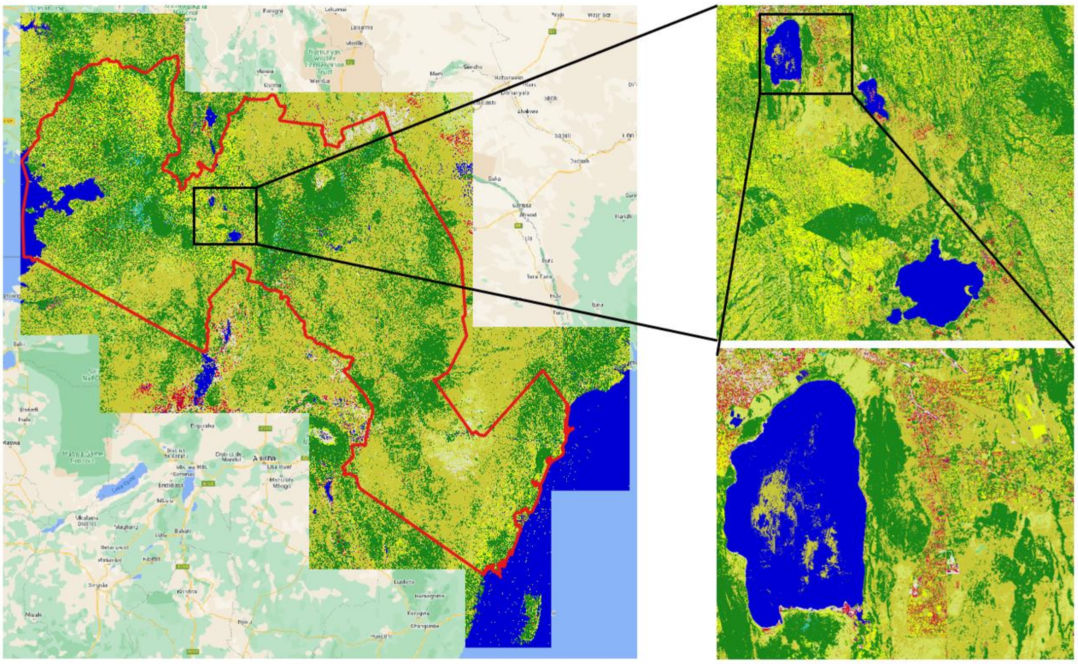


Figure 7. Raw classification output in-season crop type map Kenya

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

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 in-season crop mask is as follows:

Crop Type S2 map = (1 of 42 individual crop types or mixed cropping): Crops

Crop Type S2 map = (forest, natural shrubs, natural grassland, bare, urban, aquatic vegetation, water, wetlands): Other landcover

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 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 46 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 & 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_InSeason_LongRains_2023.tif & D3.2_ Kenya_CropMask_InSeason_LongRains_2023.tif). The nomenclature can be viewed by opening the accompanying *.lyr files 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 of the complete non-agri mask, using recent Sentinel-2 satellite data. 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.

Deviations from feasibility study proposal and the short rains season:

There were no deviations from what was described in the feasibility study and what was done during the short rains season 2022-2023.

3.3 Map production

Both the Crop Type map & Crop Mask are presented in A0 printable PDF map 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 in-season Crop Mask and Crop Type map for Kenya for the long rains season.

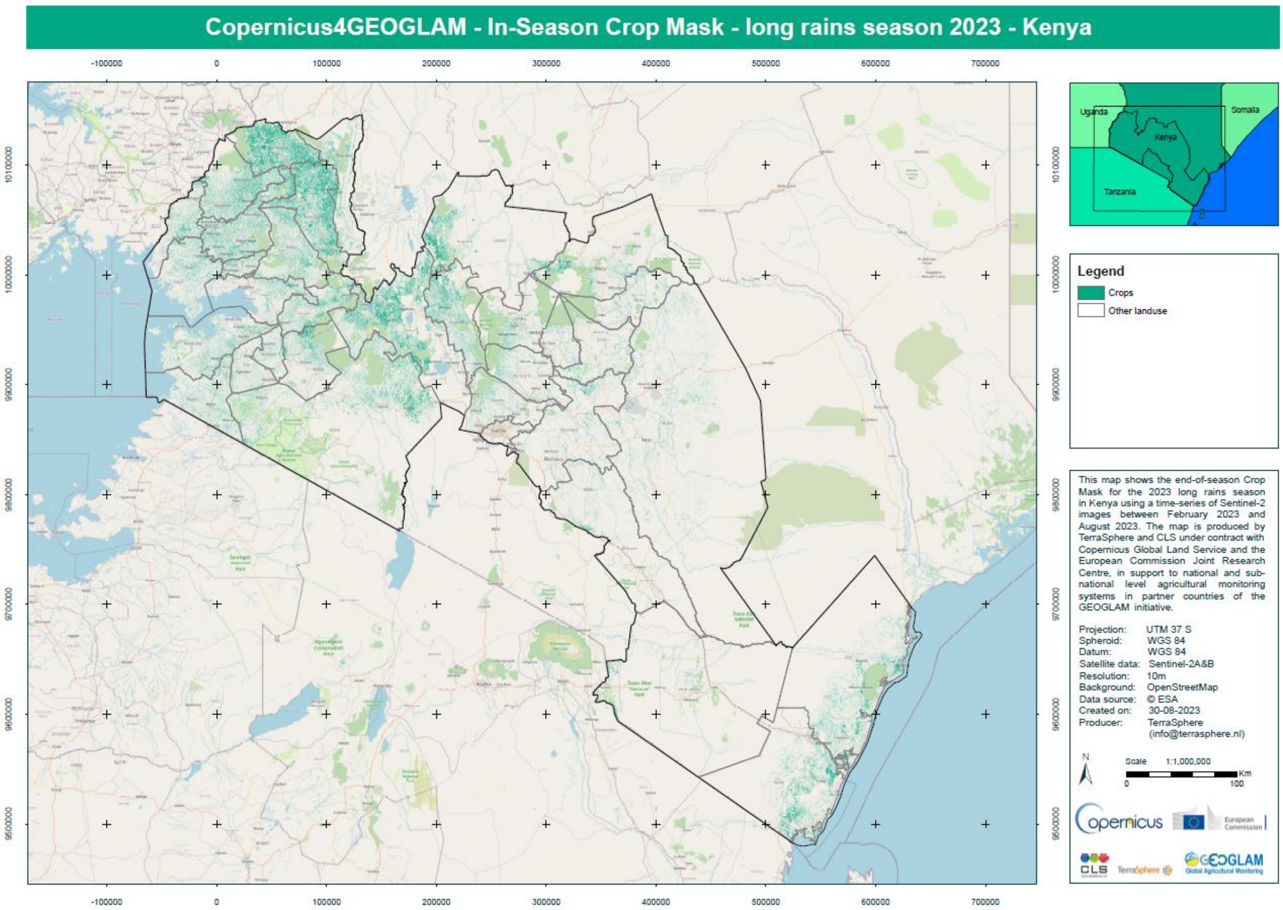


Figure 8. In-season Crop Mask for the long rains season 2023 in Kenya

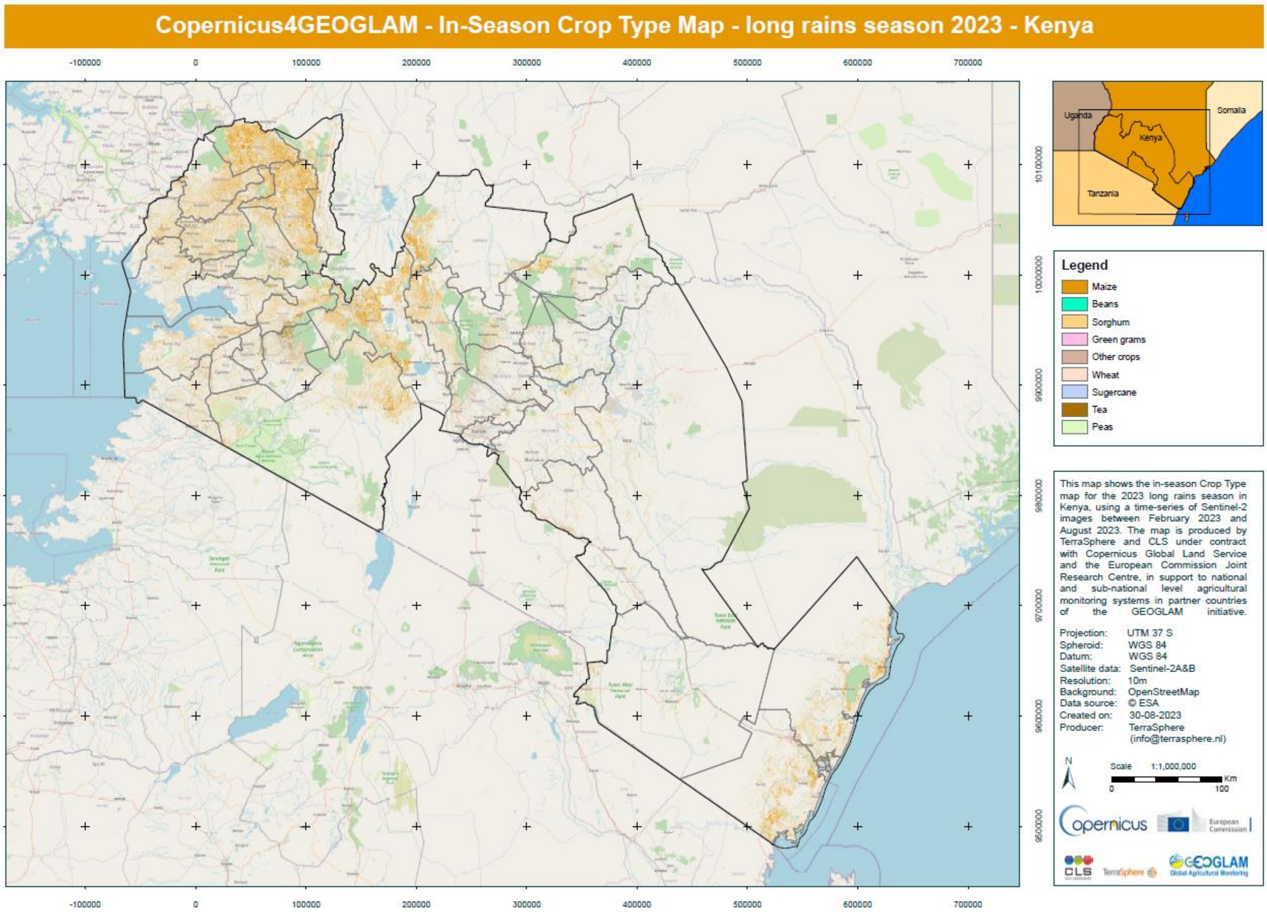


Figure 9. In-season Crop Type map for the long rains season 2023 in Kenya

Deviations from feasibility study proposal and the short rains season:

There were no deviations from what was described in the feasibility study and what was done during the short rains season 2022-2023.

3.4 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. The procedures for validation were carried out as described in the technical offer. There was no need to apply correction factors because an equal sampling intensity was applied to each stratum.

Crop type in-season mapping for Kenya (long rains)	Reference												
Map	Maize	Beans	Sorghum	Green grams	Wheat	Sugarcane	Tea	Peas	Other crops	Other landcover	Total	User Accuracy	F-Score
Maize	8111	20	23		154	441	66	129	317	2599	11860	0,68	0,44
Beans	14	6								7	27	0,22	0,05
Sorghum											0	N/A	N/A
Green grams				4						4	8	0,50	0,03
Wheat	1				193				8	4	206	0,94	0,24
Sugarcane	75					137				120	332	0,41	0,13
Tea	6						589		84	75	754	0,78	0,55
Peas	11							49		98	158	0,31	0,08
Other crops	75			24				6	1074	361	1540	0,70	0,22
Other landcover	16786	177	130	284	1059	1234	726	832	6827	243803	271858	0,90	0,94
Total	25079	203	153	312	1406	1812	1381	1016	8310	247071	286743		
Producer accuracy	0,32	0,03	0,00	0,01	0,14	0,08	0,43	0,05	0,13	0,99			

Overall accuracy	0,89
Crop Type User accuracy	0,68
Crop Type Producer accuracy	0,26

Figure 10. Confusion matrix for in-season Crop Type map of the long rains season 2023

Cropland mask in-season mapping for Kenya (long rains)

		Reference				
		Non-crop	**Crop**	**Total**	**User Accuracy**	**F-Score**
Map	**Non-crop**	**243803**	28055	**271858**	0,90	0,94
	Crop	3268	**11617**	**14885**	0,78	0,43
	Total	**247071**	**39672**	**286743**		
	Producer accuracy	0,99	0,29		**Overall accuracy**	**0,89**

Figure 11. Confusion matrix for in-season Crop Mask of the short rains season 2022-2023

Figure 10 and Figure 11 show that the overall accuracy for the Crop Type map and Crop Mask is 89% for both, higher than the specifications mentioned in the feasibility study report (D1.1) (65% for both).

The crop mask for the in-season shows satisfying results for the user accuracy of the crop class (78%) with 22% of commission errors. Nevertheless, the producer accuracy is low (29%) resulting in large omission errors (71%), which means that the classification has missed quit some cropped areas.

The omission phenomenon is also confirmed by the confusion matrix of the crop type map, which shows low producer accuracies but very satisfying user accuracies for the different crop types. The commission errors for the maize, tea, and wheat classes are respectively 32%, 22% and 6%. Overall, the wheat, maize and tea classes tend to show satisfying results with F1-Score between 0.45 and 0.55. For other classes, results are very low with F1-Score below 0.15 (beans, green grams, sugarcane 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 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.

1212**Deviations from feasibility study proposal and the short rains season:**

There were no deviations from what was described in the feasibility study and what was done during the short rains season 2022-2023.

3.5 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 in-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 13 illustrates the change with one example.

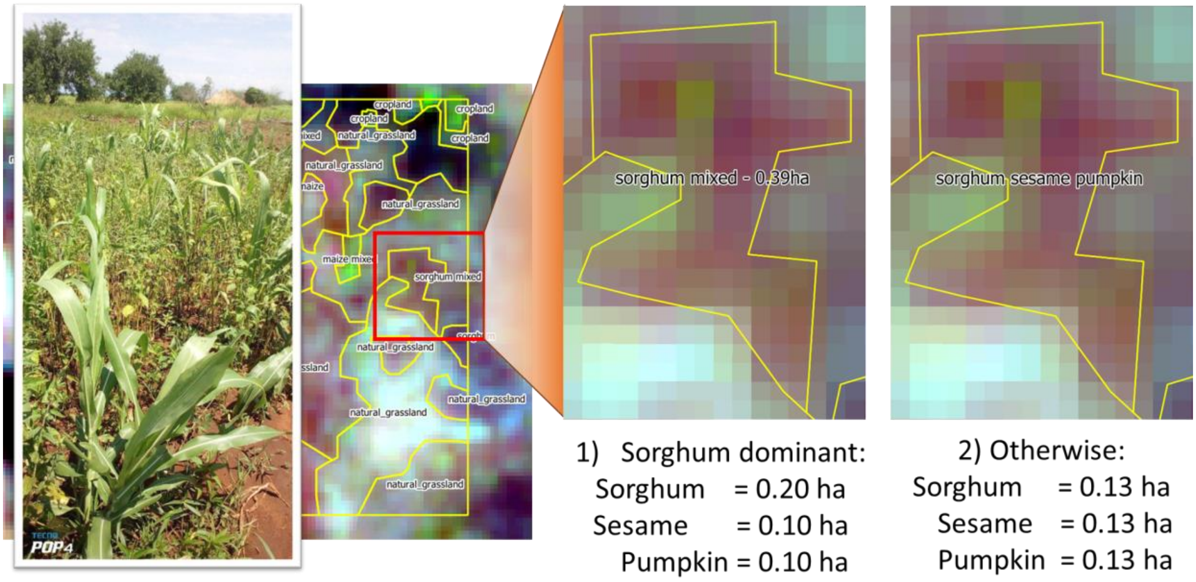


Figure 13: Mixed cropping fields and crop area estimates (non-dominant crop study case)

(2) Crop area estimates can be derived directly from the in-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 mis-classification.

(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.

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 Maize, Wheat and Tea, the same reduction in variance would have been achieved by increasing the size of the field survey sample by 3, 7 and 12.5.

Table 4: Area estimates for the in-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)	**944 405,33**	**675,80**	**36,16**	**107,87**	**7 958,80**	**11 206,35**	**63 794,25**	**1 208,10**	**31 820,22**	**17 041 220,08**
	Map (%)	0,05	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,94
Regression Estimator	Regression estimate	0,09	0,01	0,00	0,00	0,00	0,00	0,01	0,02	0,03	0,84
	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	**1 592 716,79**	**109 598,44**	**53 401,87**	**88 082,74**	**41 425,87**	**89 218,54**	**123 995,73**	**280 592,29**	**459 089,02**	**15 230 344,73**
	Variance	3 600 243 041,27	231 487 571,33	81 462 332,48	220 181 159,91	171 210 526,77	439 535 699,24	93 961 041,11	1 072 702 200,57	1 835 581 078,15	10 238 315 952,51
	Standard Error	60 002,03	15 214,72	9 025,65	14 838,50	13 084,74	20 965,11	9 693,35	32 752,13	42 843,68	101 184,56
	95% Confidence Interval	117 603,97	29 820,84	17 690,27	29 083,47	25 646,10	41 091,61	18 998,97	64 194,18	83 973,62	198 321,75
Efficiency	Regression Estimator	**3,00**	**1,12**	**1,25**	**1,28**	**6,73**	**2,45**	**12,50**	**1,29**	**1,18**	**2,17**

4 Conclusions

The overall accuracies for the long rains in-season Crop Type map and Crop Mask for the long rains are 89% for both products, which is higher than what was mentioned in the feasibility study (both 65%). Some classes such as wheat, maize and tea show satisfying and encouraging results (with F1-Score between 0.45 and 0.55) but for some individual crops (e.g. beans, green grams, sugarcane 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 by erroneous crop detection. Nevertheless, the producer accuracy of the two products is low resulting in large omission errors, which means that the classification has missed quit some cropped areas. 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.