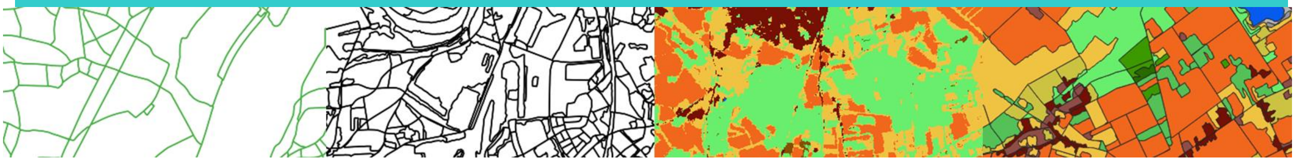
Service contract for the Copernicus Land monitoring services

Crop Mapping for GEOGLAM Country Level Support



Framework contract 939708-2020-IPR

End-of-season Crop Type Map & Crop Mask

Tanzania - long rains season - 2021

Prepared by:

CLS

COLLECTE LOCALISATION SATELLITES

&

TerraSphere

Reference: End-of-season mapping - Tanzania - long rains season - 2021

Issue 1.1 - 26/11/2021

61 rue de la Cimaise – Bâtiment C, 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

1 Introduction 1
2 Summary of data used 1
2.1 Satellite data 2
2.2 Fieldwork data 3
3 Workflow 5
3.1 Pre-processing 5
3.2 Classification 9
3.3 Map production 11
3.4 Validation 12
3.5 Area estimates 15
4 Conclusions 18

LIST OF FIGURES

Figure 1. Tanzania AOI overlaid with the S2 tile-based grid and the fieldwork segments 1
Figure 2. Sentinel-1 acquisition dates over Tanzania 3
Figure 3. Preparation of fieldwork data for training and validation 4
Figure 4. Typical cultivation in Tanzania with small plot size and multiple crops 5
Figure 5. Sentinel-1 synthetic colour composite over the Tanzanian AOI (rede outline; false colour RGB composition = Minimum, Mean, Standard Deviation) 6
Figure 6: Sentinel-1 monthly synthesis images for (S2) tile 3MYA. Image dates ranging from January (upper left) until July (lower left) 7
Figure 7: Sentinel-2 monthly synthesis composite (true colour RGB composition), 15/03/2021, tile 36MZU 8
Figure 8. Raw classification output end-of-season crop type map Tanzania 9
Figure 9. End-of-season Crop Mask for Tanzania 2021 11
Figure 10. End-of-season Crop Type map for Tanzania 2021 12
Figure 11. Confusion matrix for end-of-season Crop Type map 13
Figure 12. Confusion matrix for end-of-season Crop Mask 14
Figure 13: Mixed cropping fields and crop area estimates 15

LIST OF TABLES

Table 1. S2 tiles covering the AOI for Tanzania 2
Table 2. Nomenclature for Crop Mask 10
Table 3. Nomenclature for Crop Type map 10
Table 4: Area estimates for the end-of-season mapping 17

Issues :

1.0	08-11-2021	First version submitted to JRC
1.1	26-11-2021	Correction of minor issue in Table 4

1 Introduction

This document describes the end-of-season mapping of the crop type and crop mask for the Area Of Interest (AOI) in Tanzania. It 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 in-season mapping. The document describes also the satellite imagery and the ground truth data actually 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 Tanzania.

2 Summary of data used

The Figure 1 shows the AOI for Tanzania, the grid of all Sentinel-2 tiles and the fieldwork (500x500m) segments.

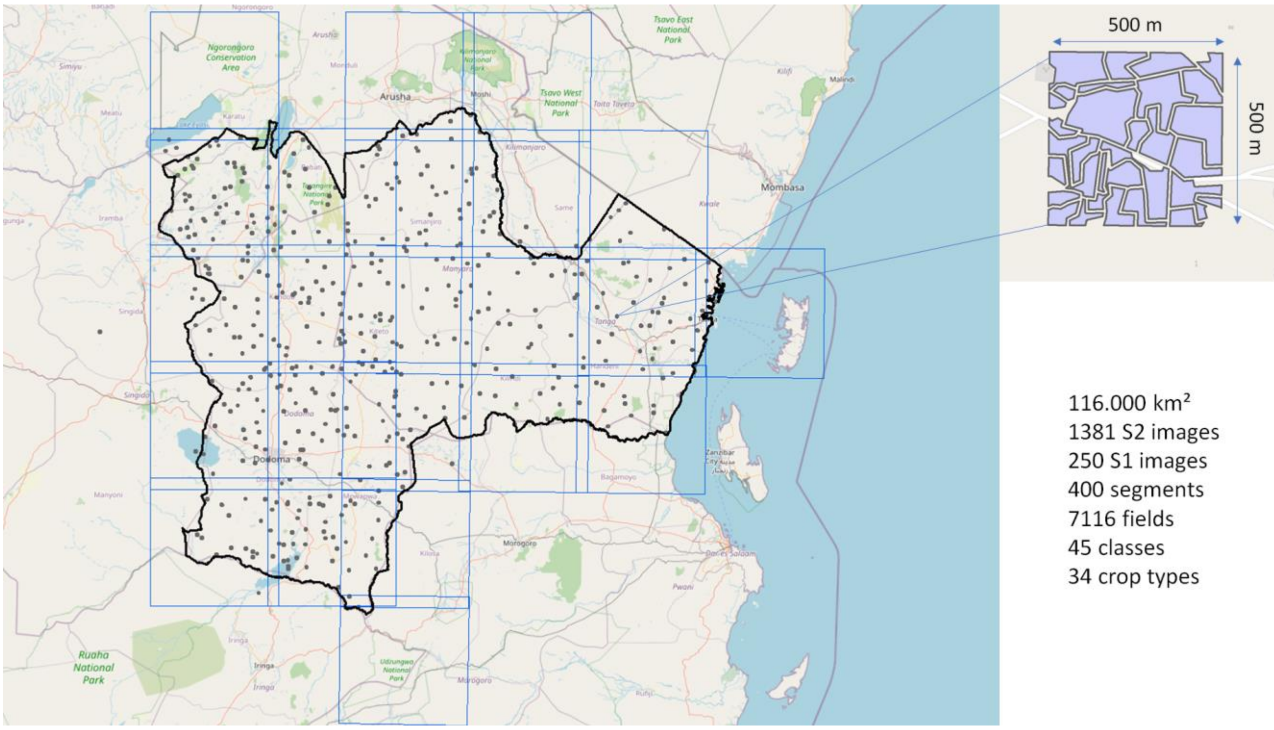


Figure 1. Tanzania AOI overlaid with the S2 tile-based grid and the fieldwork segments

2.1 Satellite data

Sentinel-2

In total, 1,381 Sentinel-2A & B Level-2A images have been acquired covering 23 tiles between 02-10-2020 and 17-09-2021 for the end-of-season mapping. Early dates are only relevant for Dodoma which experiences earlier planting due to the unimodal season. For the other two regions (Tanga and Manyara), only dates from 01-02-21 are relevant. The Table 1 lists the S2 data used per S2 tile ID.

Table 1. S2 tiles covering the AOI for Tanzania

Tile ID	First Date	Last Date	Number of Images
36MYA	03/10/2020	10/09/2021	107
36MYT	03/10/2020	10/09/2021	91
36MYU	03/10/2020	08/09/2021	100
36MYV	03/10/2020	31/08/2021	87
36MZA	05/10/2020	31/08/2021	29
36MZB	03/10/2020	26/08/2021	85
36MZT	05/10/2020	12/07/2021	42
36MZU	05/10/2020	31/08/2021	44
36MZV	05/10/2020	31/08/2021	37
37MBM	02/10/2020	09/07/2021	65
37MBN	02/10/2020	31/08/2021	71
37MBP	05/10/2020	31/08/2021	37
37MBQ	05/10/2020	31/08/2021	34
37MBR	05/10/2020	31/08/2021	31
37MBS	05/10/2020	10/09/2021	40
37MCP	02/10/2020	28/08/2021	62
37MCQ	02/10/2020	31/08/2021	65
37MCR	02/10/2020	28/08/2021	81
37MCS	02/10/2020	31/08/2021	74
37MDP	02/10/2020	07/09/2021	54
37MDQ	02/10/2020	28/08/2021	45
37MDR	02/10/2020	17/09/2021	41
37MEQ	02/10/2020	07/09/2021	59

Sentinel-1

In total, approximatively 250 Sentinel-1 images have been used to cover the Tanzanian AOI between 01-01-2021 and 29-08-2021 for the end-of-season mapping. The Figure 2 shows the acquisition dates of the S1 dataset covering the Tanzanian AOI. The coverage consists of three descending (southward) orbits, requiring 3 to 4 images per orbit.

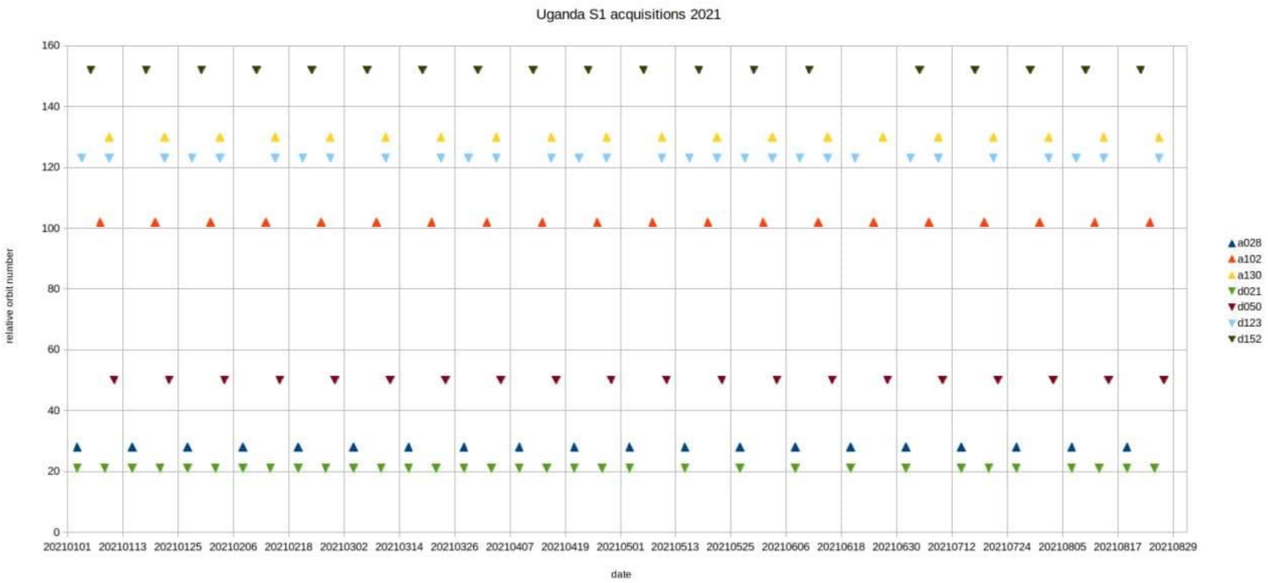


Figure 2. Sentinel-1 acquisition dates over Tanzania

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 0.2 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.

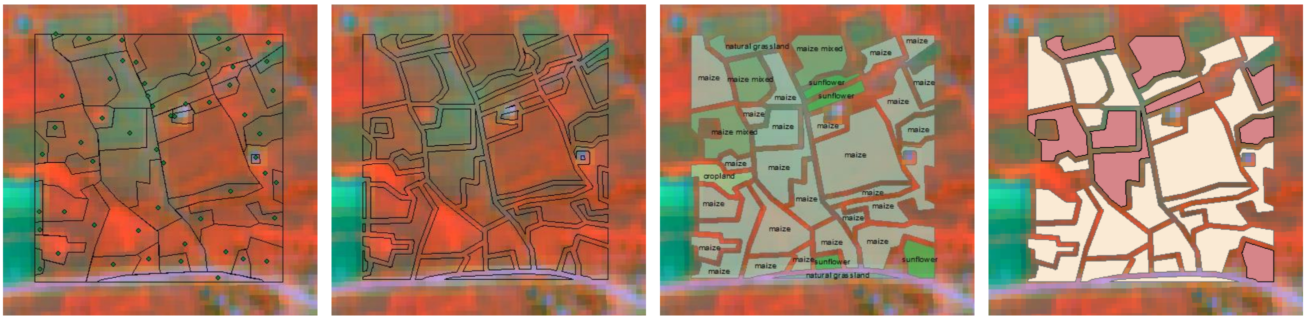
2) A negative buffer of 5 meters is applied to eliminate, or at least minimize, the boundary effects between different classes that will negatively impact the purity of training samples signatures. Consequently, polygons are always separated by 10 meters, which corresponds to the size of 1 Sentinel-2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

3) The acreage of each buffered polygon is calculated and all polygons smaller than 0.2 ha are deleted. It should be noticed that the threshold of 0.2 ha (approximatively 18 contiguous S2 pixels) is larger than the Minimum Mapping Unit from the technical specifications and the MMU applied for the in-season mapping for the step 3 to better consider the observed agricultural practises in Tanzania. Indeed, based on the in-season mapping experience, polygons below 0.2 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) The resulting dataset from step 1 to 3 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 3.3, the classification workflow is applied per S2-based block.

5) All the resulting polygons have been visually checked and manually edited to correct obvious errors.

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 (0.2 ha)

Split between training (yellow) & validation (red)

Figure 3. Preparation of fieldwork data for training and validation

Resulting from all the described processing steps, 3,775 polygons, covering approximatively 9,730 ha are available for the classification process. 2,950 are used for training and 825 for validation. In total 45 individual classes are distinguished, mostly individual crops (34).

Figure 4 shows an example of a typical field visited during the campaign, highlighting the small size of the plots, as well as the heterogeneity of the cultivation.

Figure 4. Typical cultivation in Tanzania with small plot size and multiple crops

Summary of the deviations from the in-season mapping:

- Deletion of polygons smaller than 0.2 ha (0.1 ha during the in-season mapping).

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) and done during the in-season mapping. For each of the two satellite data types some specific pre-processing is summarised as follows below.

Sentinel-1

Sentinel-1 Gamma0 workflow starts with Sentinel-1 level 1.1 (SLC) data products. The following steps are executed:

1. Querying Sentinel-1 repository for images acquired over area-of-interest;
2. Preparation of CopDEM 30" DEM for area-of-interest;
3. Update of local SNAP Restituted Orbit (RESORB) repository;
4. Reading S1 SLC data product: get image and metadata;
5. Applying restituted orbit file (RESORB), for improved geocoding accuracy, almost as good as Precise Orbits (PREORB) but available just after reception of the image data;
6. Thermal Noise Removal, mostly for suppressing noise patterns over large water bodies;
7. Radiometric Calibration: convert digital numbers to calibrated Gamma0 backscatter intensity values;
8. Multi-looking: combine pixels into more or less square pixels and reduce speckle noise;
9. Speckle filtering (Refined Lee) for more reduction of speckle;
10. Terrain Correction: geometric terrain correction and map projection to a 10x10m pixel grid;

11. Radiometric Terrain Correction or Slope Correction and normalization of incidence angle: dedicated script for reducing slope illumination effects using local and global incidence angle information[1];
12. Conversion from intensity values to decibel [dB] values;
13. Export to deflate-compressed geotiff file;
14. Calculation of multi-temporal statistic parameters (minimum, maximum, mean, standard deviation) over the present growing season or defined period of time;
15. Scaling and output of multi-temporal statistics to 8-bits values;
16. Storing output products in <country>/<rel.orbit>/<product> directory structure;

The Figure 5 shows a colour composite of the Sentinel-1 minimum, mean and standard deviation of images taken between January and August 2021 (ascending orbit). Yellow is seasonal stable medium-high backscatter (e.g. forest, natural shrubs, natural grassland), black is seasonal stable low backscatter (water), blue is seasonal dynamic backscatter (agriculture with the exception of the high mountainous areas which can be easily identified based on ancillary data).

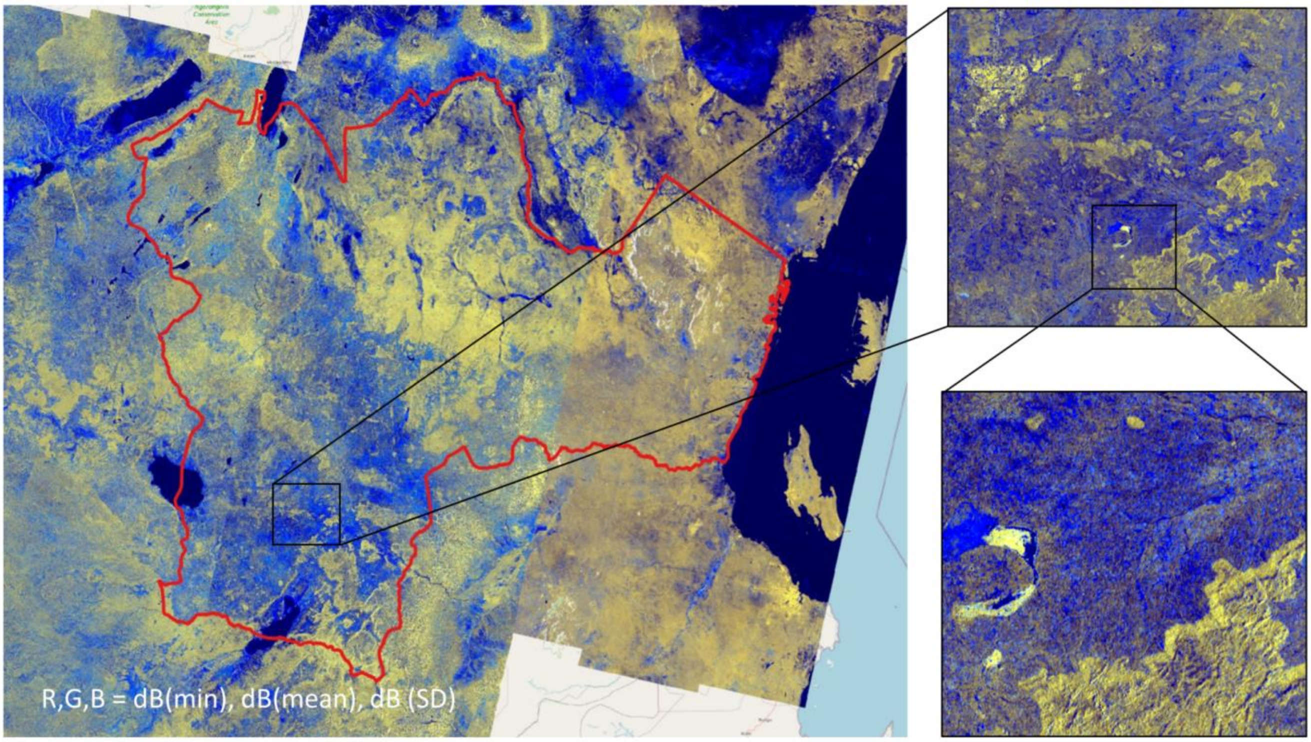


Figure 5. Sentinel-1 synthetic colour composite over the Tanzanian AOI (rede outline; false colour RGB composition = Minimum, Mean, Standard Deviation)

[1] Hoekman, D.,H., Reiche, J. Multi-model radiometric slope correction of SAR images of complex terrain using a two-stage semi-empirical approach, in Remote Sensing of Environment, 2015, doi:10.1016/j.rse.2014.08.037

Additional to the pre-processing described above, a new workflow has been set-up to produce monthly S1 synthesis images parallel to the S2 L3A data. The objective is to perform test for the crop type classification using S1 scenes or a combination between S1 and S2 images. Based on the S2 tiling grid, monthly S1 composites are produced using the median backscatter value of a period of 30 days for both VV and VH channels. S1 synthesis images are produced for January until July. The figure below shows all 7 S1 synthesis images (RGB = VV,VH,VV-VH).

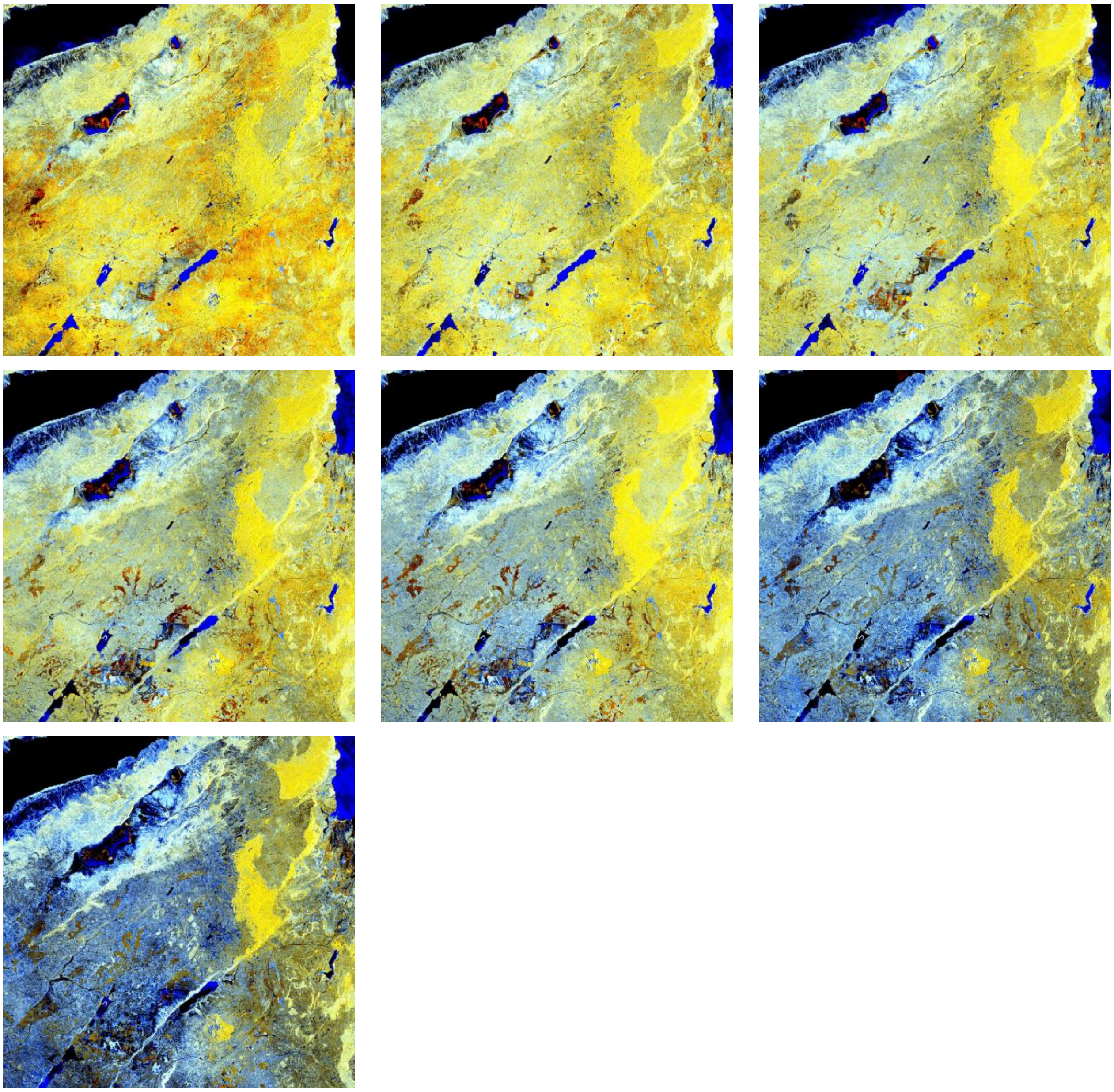


Figure 6: Sentinel-1 monthly synthesis images for (S2) tile 3MYA. Image dates ranging from January (upper left) until July (lower left)

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 7 shows an example for tile 36MZU, with a centre-date of 15-03-2021. For this synthesis, the algorithm considers all images +/- 45 days from 15-03-2021, and takes the cloud-free pixel closest to the centre date.

Figure 7: Sentinel-2 monthly synthesis composite (true colour RGB composition), 15/03/2021, tile 36MZU

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 very suitable in case of Tanzania when reviewing the size of the agricultural fields.

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

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

Summary of the deviations from the feasibility study report:

- Sentinel-2 data has been processed to monthly L3A synthesis images covering +/-45 days whereby mostly cloud-free monthly data has been obtained.
- Sentinel-1 data has been processed to synthetic channels of minimum, mean and standard deviation (Sigma nought, Db) of a seasonal stack of VH images.
- No Landsat-8 data was used for the end-of-season mapping as enough Sentinel-2 data with higher resolution was available thanks to the L3A processing.

Deviations from the in-season mapping:

Sentinel-1 monthly synthesis images based on the S2 tiling grid have been produced.

3.2 Classification

Crop Type – It was decided to take profit of the run already conducted for the in-season mapping to select the classification algorithm in Tanzania for the end-of-season. Various algorithms were tested, including supervised (maximum likelihood) classification, TempCNN and Random Forest (RF) algorithms. Based on the validation results for Tanzania, it was decided to use the RF classification as final method for the end-of-season mapping too. 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 45 classes (of which 34 crop types). The Figure 8 shows the result of the raw classification output, before post-processing.

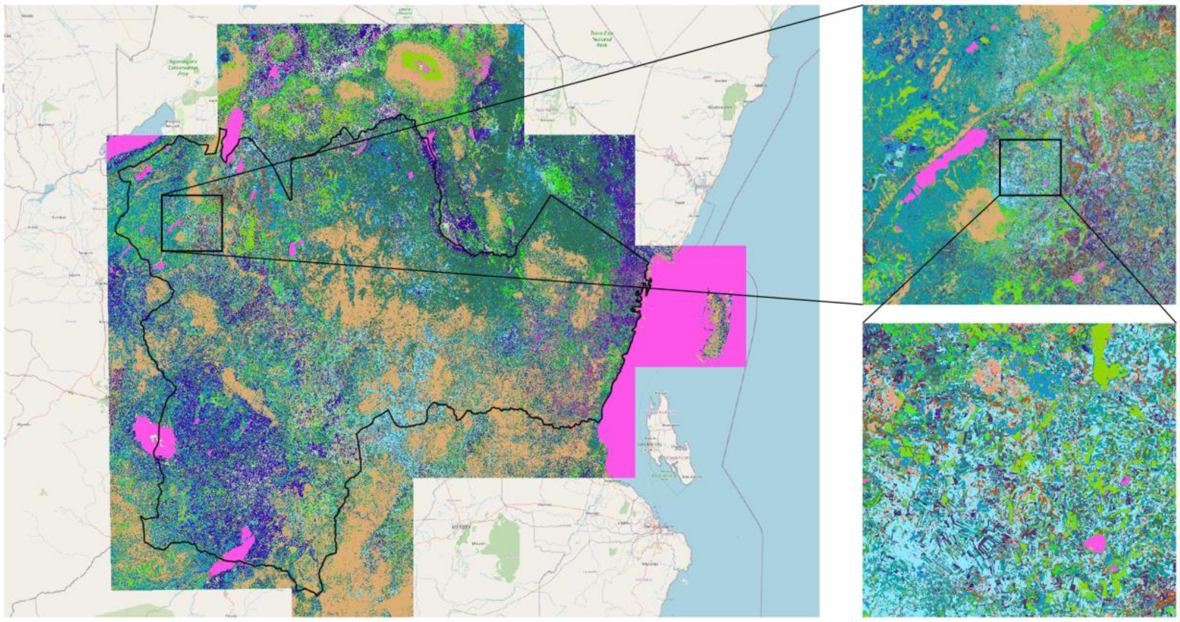


Figure 8. Raw classification output end-of-season crop type map Tanzania

For the end-of-season mapping, a test is being performed on using S1 monthly synthesis data (as well as a combination of S1 and S2) on a single tile (36MYA). Based on monthly synthesis Sentinel-2 images (L3A) and monthly Sentinal-1 data, the RF and TempCNN classifier are being tested to produce the crop type map [*at the time of writing this report, the results are not available yet, but will be reported in a new version of the document*].

Crop Mask – For the crop mask, different methods using both S2 and S1 data have been tested. A full crop mask using only Sentinel-1 data has been generated, and in parallel a full crop mask using the aggregated results from the S2-derived crop type map has been produced. Both methods yielded good results (70% for S1; 82% for S2), but since the accuracy of the S2-derived map was significantly higher, it was decided to use the S2-based product as the final Crop Mask for the end-of-season map. The rule to produce the current end-of-season crop mask is as follows:

Crop Type S2 map = (1 of 34 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 results of the test to use a combination of S1 and S2 data as described above for mapping crop types will also be used to analyse if we can further increase Crop Mask accuracy. 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, 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 45 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 Table 3 lists the final classes for the Crop Type map and number coding as found in the final GeoTiff files (Tanzania_CropType_EndOfSeason_LongRainy_2021.tif & Tanzania_CropMask_EndOfSeason_LongRainy_2021.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
2	sunflower	
3	millet	
4	beans	
5	sorghum	
6	green grams	
7	sisal	
8	cassava	
9	other crops	all other monoculture crops and mixed cropping
10	other landcover	Forest, water, natural shrubs, natural grassland, urban, bare, aquatic vegetation, wetlands

Some obvious classification errors have been recoded, e.g. the presence of crops in large water bodies. A shapefile on protected area boundaries was used to recode erroneous cropland to other landcover, as no agriculture is legally supposed to be taking place in these areas. However, agricultural encroachment may sometimes take place in these protected areas (rare) and they were preserved in the final map. As a final step a majority filtering have been applied using a moving box-size of 3x3 pixels (= approximate MMU for S2), as well as a “sieve” operation whereby all pixel clusters of 4 pixels (0.04ha, according to technical specification) are recoded to the majority surrounding class. All maps are presented in UTM, zone 37 South.

Deviations from feasibility study proposal and the in-season mapping:

There’s been no substantial deviations from what has been described in the feasibility study and done for the in-season mapping.

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 1:600.000 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 Tanzania.

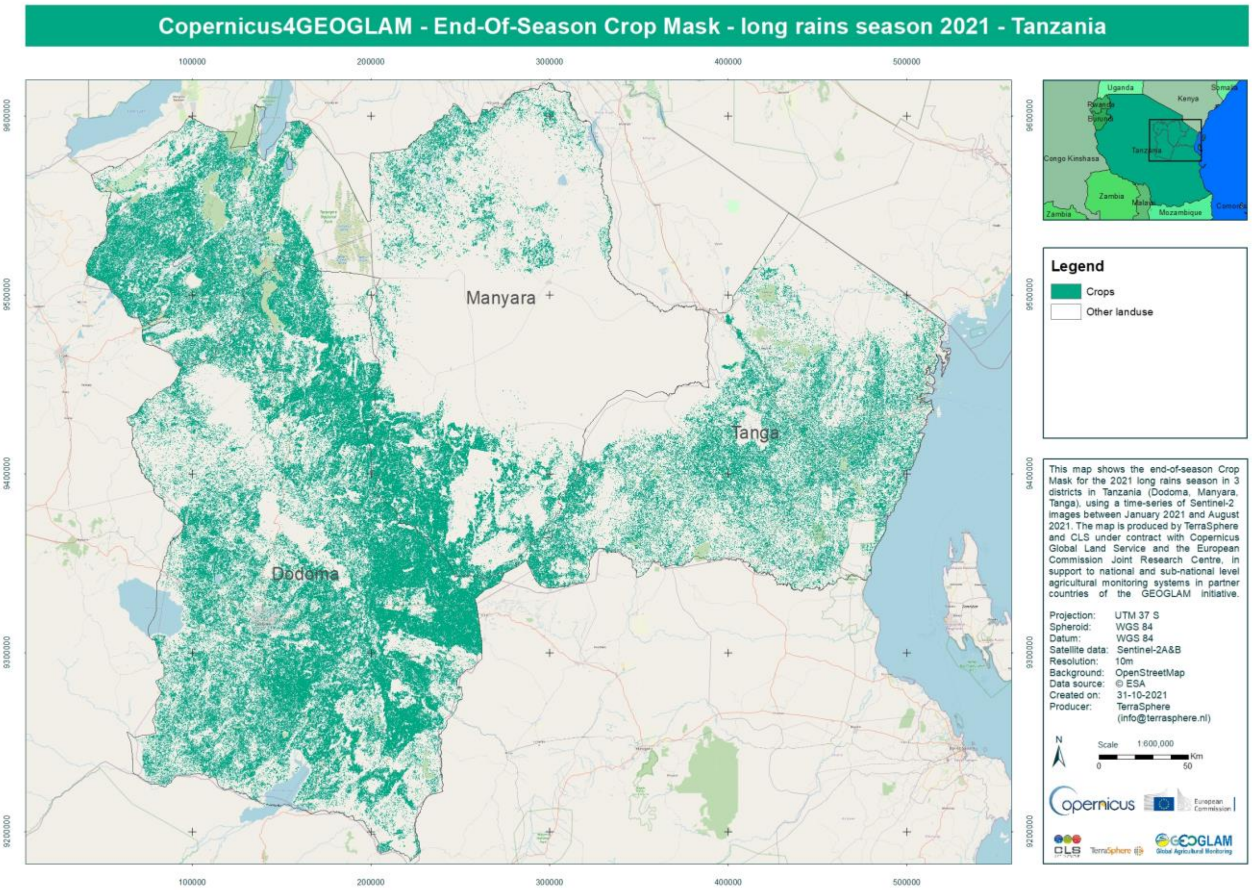


Figure 9. End-of-season Crop Mask for Tanzania 2021

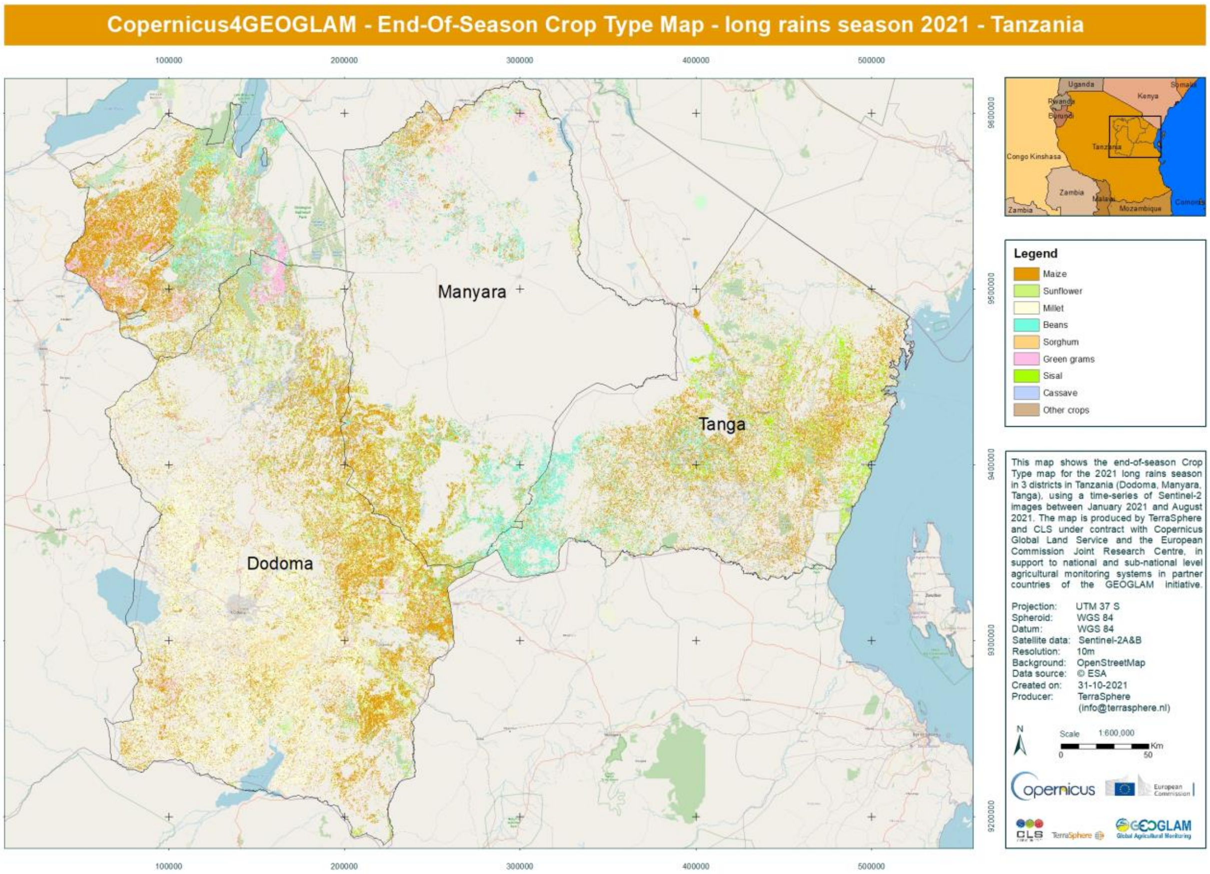


Figure 10. End-of-season Crop Type map for Tanzania 2021

Deviations from feasibility study proposal and the in-season mapping:

There's been no substantial deviations from what has been described in the feasibility study and done for the in-season mapping.

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 Figure 11 and Figure 12 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 end-of-season mapping for Tanzania	Reference												
Map	Maize	Sunflower	Millet	Beans	Sorghum	Green grams	Sisal	Cassava	Other crops	Other landcover	Total	User Accuracy	F-Score
Maize	**15272**	1012	133	21	133	23		43	125	4075	**20837**	0,73	0,59
Sunflower	1171	**2529**	38		37				14	1838	**5627**	0,45	0,47
Millet	2230	549	**1273**		414			14	25	5175	**9680**	0,13	0,22
Beans	1129	161		**954**				13	26	1810	**4093**	0,23	0,35
Sorghum	828	120	259	1	**888**					665	**2761**	0,32	0,40
Green grams	484	12	18	11	11	**1400**			6	519	**2461**	0,57	0,71
Sisal	102						**847**	7	4	600	**1560**	0,54	0,70
Cassava	1273	143	132	16	18			**313**		1238	**3133**	0,10	0,17
Other crops	1045	125	6			35			**869**	1101	**3181**	0,27	0,38
Other landcover	7350	595	239	299	180	21	4	100	295	**136387**	**145470**	0,94	0,91
Total	**30884**	**5246**	**2098**	**1302**	**1681**	**1479**	**851**	**490**	**1364**	**153408**	**198803**		
Producer accuracy	0,49	0,48	0,61	0,73	0,53	0,95	1,00	0,64	0,64	0,89			

Overall accuracy	0,81
Crop Type User accuracy	0,46
Crop Type Producer accuracy	0,54

Figure 11. Confusion matrix for end-of-season Crop Type map

Cropland mask end-of-season mapping for Tanzania		Reference				
		Non-crop	Crop	Total	User Accuracy	F-Score
Map	Non-crop	136387	9083	145470	0,94	0,91
	Crop	17021	36312	53333	0,68	0,74
	Total	153408	45395	198803		
	Producer accuracy	0,89	0,80			
					Overall accuracy	0,87

Figure 12. Confusion matrix for end-of-season Crop Mask

Figure 11 and Figure 12 show that the overall accuracy for the Crop Type map and Crop Mask is respectively reaching 81% and 87%, which is greater than the specifications mentioned in the feasibility study report (D1.1) (65% & 65%). The crop mask for the end-of-season shows very satisfying results for both user and producer accuracies for the crop class with respectively 32% and 20% commission and omission errors (Commission = 100% - User Accuracy, Omission = 100% - Producer Accuracy).

Large improvements of the accuracies are also to be noted compared to the in-season mapping (respectively 71% and 82% for the crop mask and crop type mapping) due to the addition of the end-of-season imagery and the selection of training polygons greater than 0.2 ha. The end-of-season mapping massively reduces both omission and commissions errors for the crop classes (approximatively 20% reduction of the errors). The results for individual crops reach very satisfying accuracies with F-Score nearly greater than 0.6 for the classes “Maize”, “Green grams” and “Sisal”. F-Score increase by nearly 0.2-0.3 for many classes such as “Maize” (from 0.37 to 0.59), “Sunflower” (from 0.23 to 0.47), “Sorghum” (from 0.14 to 0.40), “Green grams” (from 0.43 to 0.71) and “Other crops” (from 0.11 to 0.38). Nevertheless, going beyond the requirements and improvements, some accuracies focusing on the crop type remains weak for both user and producer results. The classes “Green grams” and “Sisal” show the best individual results with F1-Score greater than 0.7. The lower results are obtained for the classes “Cassava” and “Millet”.

Deviations from feasibility study proposal and the in-season mapping:

There’s been no substantial deviations from what has been described in the feasibility study and done for the in-season mapping.

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 end-of-season map alone;
3. Regression estimators: area estimates derived from field data combined with end-of-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.

Some deviations have been applied compared to the in-season mapping to better consider the mixed cropping practice. For the in-season area estimates, only the dominant crop was considered for the mixed cropping parcels whereas for the end-of-season mapping, all the crop surveyed in the field were taking into account for the estimates, contributing equally to the total area of the field. Figure 13 illustrates the change with one example.

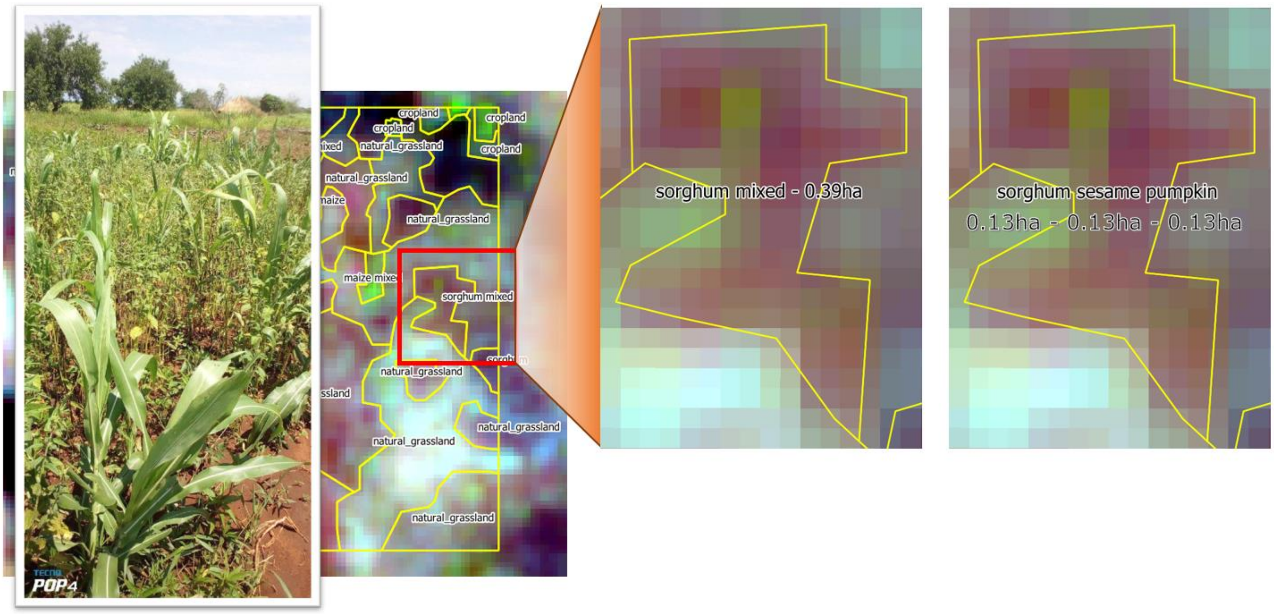


Figure 13: Mixed cropping fields and crop area estimates

(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 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 Tanzania for the end-of-season. It is interesting to notice the good relative efficiencies for maize, sunflower, millet, beans, sorghum, green grams or sisal with figures greater than 2. For example for sorghum, the same reduction in variance would have been achieved by increasing the size of the field survey sample by nearly 4.

Deviations from feasibility study proposal and the in-season mapping:

Following the feedback from the end users, some improvements have been applied compared to the in-season mapping to better consider the mixed cropping practice. For the end-of-season mapping, all the crops surveyed in a parcel are now considering for the crop area estimates, contributing equally to the total area of the parcel. Previously, only the dominant crop was considered for the associated parcel.

Table 4: Area estimates for the end-of-season mapping

		Maize	Sunflower	Millet	Beans	Sorghum	Green grams	Sisal	Cassava	Other crops	Other landcover
AOI Area (ha)	11 618 962,02										
Direct Expansion	Estimate of proportion	0,11	0,06	0,02	0,02	0,01	0,01	0,01	0,01	0,05	0,71
	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	0,00
	95% Confidence Interval	0,02	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	Estimate of the class area	**1 297 793,59**	**704 541,17**	**194 828,24**	**192 032,42**	**156 824,74**	**100 614,98**	**82 325,59**	**115 922,23**	**574 028,51**	**8 200 050,54**
	Variance	12 483 972 992,08	4 637 155 881,44	1 265 251 638,57	2 008 168 975,53	1 065 098 741,89	2 050 377 483,54	2 293 096 530,81	757 166 312,63	5 345 915 464,25	45 300 671 197,61
	Standard Error	111 731,70	68 096,67	35 570,38	44 812,60	32 635,85	45 281,09	47 886,29	27 516,66	73 115,77	212 839,54
	95% Confidence Interval	218 994,13	133 469,46	69 717,94	87 832,69	63 966,27	88 750,94	93 857,12	53 932,64	143 306,90	417 165,50
Pixel count	Map (ha)	**1 279 253,34**	**336 968,60**	**715 332,06**	**353 913,57**	**174 552,27**	**163 642,71**	**167 254,48**	**300 634,41**	**148 241,11**	**7 979 169,48**
	Map (%)	0,11	0,03	0,06	0,03	0,02	0,01	0,01	0,03	0,01	0,69
Regression Estimator	Regression estimate	0,09	0,05	0,02	0,02	0,01	0,01	0,01	0,01	0,05	0,73
	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	0,00
	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	**1 069 686,88**	**633 446,36**	**215 404,75**	**223 927,21**	**139 586,36**	**81 966,04**	**74 424,83**	**128 329,97**	**564 557,79**	**8 429 373,80**
	Variance	4 231 662 105,79	2 051 232 241,15	624 105 993,20	953 345 124,38	267 114 521,85	628 561 173,62	433 645 420,63	541 799 766,79	4 362 307 228,50	15 584 161 089,59
	Standard Error	65 051,23	45 290,53	24 982,11	30 876,29	16 343,64	25 071,12	20 824,15	23 276,59	66 047,76	124 836,54
	95% Confidence Interval	127 500,40	88 769,44	48 964,94	60 517,52	32 033,53	49 139,40	40 815,34	45 622,12	129 453,62	244 679,61
Efficiency	Regression Estimator	**2,95**	**2,26**	**2,03**	**2,11**	**3,99**	**3,26**	**5,29**	**1,40**	**1,23**	**2,91**

4 Conclusions

The availability of cloud-free Sentinel-2 data over the AOI was lower than expected during the feasibility study based on 5-yearly cloud statistics. However, the processing to L3A 45-day synthesis yields very good results and creates monthly nearly cloud-free data tiles with which crop classification is feasible. Various crop type classification methods have been tested of which RF yields the best results. The overall accuracy for the end-of-season Crop Type map is 81% and the end-of-season Crop Mask 87%, which is better than what was mentioned in the feasibility study (both 65%). The end-of-season results are also better than the in-season results due to the addition of the end-of-season imagery and the selection of training polygons greater than 0.2 ha. For some individual crops though (e.g. beans and groundnuts), lower accuracies are still reported. Tests using a combination of S1 and S2 are conducted to improve the end-of-season mapping by including SAR data in the classification algorithm [*results not available at the time of writing the version 1.0 of the report*].