

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 - short rains season – 2022-2023

Prepared by:



CLS
COLLECTE LOCALISATION SATELLITES

&

TerraSphere 

Reference : End-of-season mapping - Tanzania - short rains season – 2022-2023

Issue 1.0 - 17/05/2023

Limited distribution/Diffusion limitée

61 rue de la Cimaïse, Bâtiment C, Green Park, 59650 Villeneuve d'Ascq, France

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

www.cls.fr

TABLE OF CONTENTS

1	Introduction	1
2	Summary of data used	2
2.1	Satellite data – Sentinel-2	2
2.2	Fieldwork data.....	3
3	Workflow.....	7
3.1	Pre-processing.....	7
3.2	Classification.....	9
3.3	Map production.....	11
3.4	Validation	13
3.5	Area estimates.....	15
4	Conclusions	18

LIST OF FIGURES

Figure 1: Warning map - End of March (Source: ASAP - Anomaly Hotspots of Agricultural Production)	1
Figure 2. Tanzania AOI overlaid with the S2 tile-based grid and the fieldwork segments	2
Figure 3. Preparation of fieldwork data for training and validation.	4
Figure 4. Field with sisal & maize in mixed cropping	5
Figure 5. Field at an early stage of growth	6
Figure 6. Bare field prepared for cultivation but awaiting (belated) rainfall	7
Figure 7: Sentinel-2 monthly synthesis composite (infrared false color composite), 15/02/2023, tile 36MXA.	8
Figure 8. Raw classification output end-of-season crop type map Tanzania	9
Figure 9. End-of-season Crop Mask for the short rains season 2022-2023 in Tanzania	11
Figure 10. End-of-season Crop Type map for the short rains season 2022-2023 in Tanzania	12
Figure 10. Confusion matrix for end-of-season Crop Type map of the short rains season 2022-2023	13
Figure 11. Confusion matrix for end-of-season Crop Mask of the short rains season 2022-2023	14
Figure 14: Mixed cropping fields and crop area estimates (non-dominant crop study case)	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 of the short rains season 2022-2023 in Tanzania	17

Issues :

1.0	17-05-2023	First version submitted to JRC
-----	------------	--------------------------------

1 Introduction

This document describes the end-of-season mapping of the crop type and crop mask for an Area Of Interest (AOI) in Tanzania during the short rains season 2022-2023 covering the region of Tanga and Manyara. 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 Tanzania.

The current short rains season in Tanzania has been affected by a severe drought¹², that has affected the - eastern regions of Tanzania (including both regions Tanga and Manyara) but the results of the end-of-season mapping don't seem affected. The drought has probably affected crop growth during the season in these regions. Crop and livestock production has been severely affected or failed in large parts of the Horn (mainly in Somalia, but also areas of Kenya, southern Ethiopia and parts of Uganda and Tanzania). This region has experienced a negative record of five consecutive failed rainy seasons, causing major concern to food availability and access. March has seen heavy rains in the Horn of Africa due to atmospheric humidity linked to the cyclones in the Mozambique corridor. In addition, seasonal precipitation forecasts for April-May have improved compared with earlier forecasts. The rains will favour planting and early-stage crop development. However, due to prolonged consecutive droughts, many livelihoods have been eroded and it will take some time for them to fully recover. Due to the severe and prolonged drought that has affected parts of Ethiopia, Kenya, Somalia and to a lesser extent Uganda and Tanzania, overall vegetation conditions are still below average for this time of year. Other sources say that weather forecasts suggest that there is a reasonable chance that the upcoming March-May rains will again be below average³. If this happens, it will be an unprecedented sixth poor season, with catastrophic humanitarian consequences.

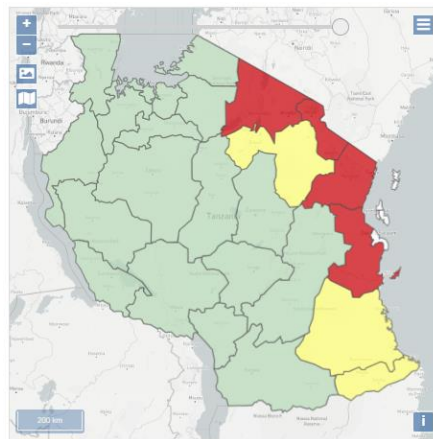


Figure 1: Warning map - End of March (Source: ASAP - Anomaly Hotspots of Agricultural Production)

¹ <https://mars.jrc.ec.europa.eu/asap/country.php?cntry=257>

² <https://droughtwatch.icpac.net/mapviewer/>

³ <https://www.rescue.org/press-release/irc-severe-drought-projected-leave-about-54-million-people-kenya-without-adequate>

2 Summary of data used

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

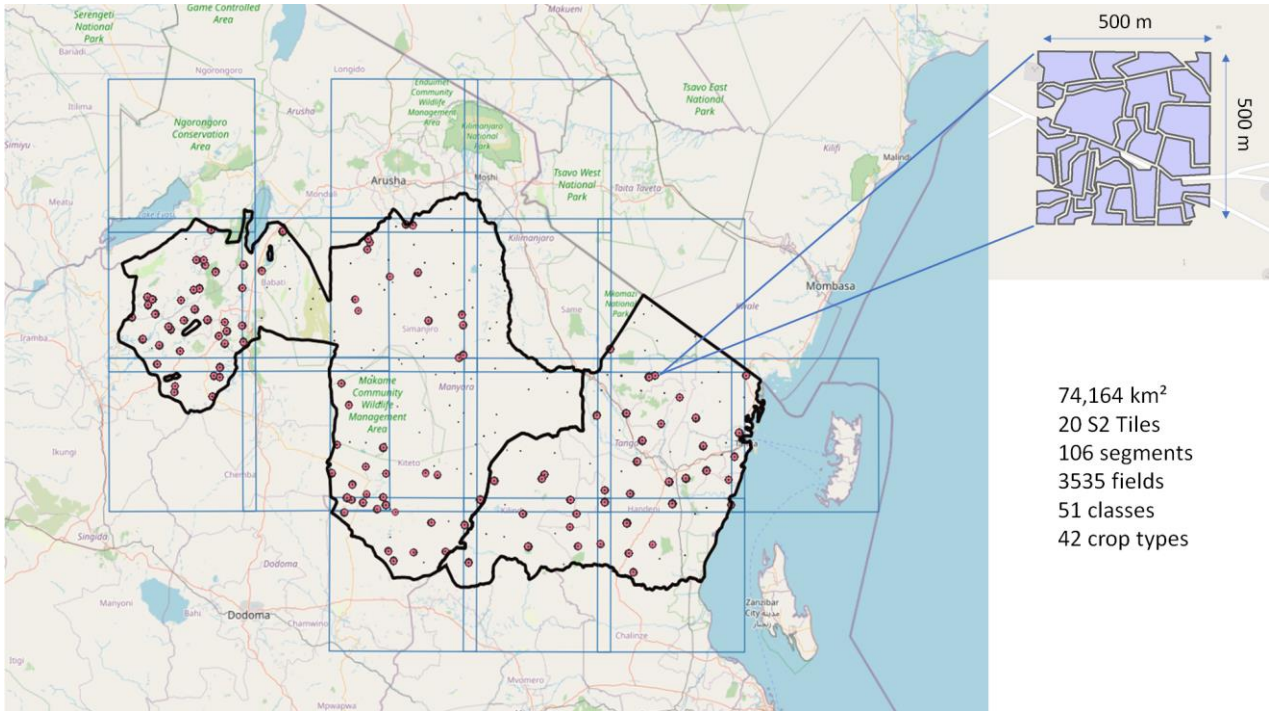


Figure 2. Tanzania AOI overlaid with the S2 tile-based grid and the fieldwork segments

2.1 Satellite data – Sentinel-2

In total, approximately 797 Sentinel-2A & B Level-2A images have been acquired covering 20 tiles between 02-09-2022 and 27-03-2023. 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
36MXA	03/09/2022	27/03/2023	33
36MYA	03/09/2022	27/03/2023	61
36MYB	03/09/2022	22/03/2023	69
36MYV	03/09/2022	27/03/2023	55
36MZA	05/09/2022	09/03/2023	24
36MZB	03/09/2022	27/03/2023	64
36MZU	05/09/2022	22/02/2023	25
36MZV	05/09/2022	09/03/2023	24
37MBP	05/09/2022	22/02/2023	24

Tile ID	First Date	Last Date	Number of Images
37MBQ	05/09/2022	24/03/2023	26
37MBR	05/09/2022	09/03/2023	24
37MBS	05/09/2022	14/03/2023	25
37MCP	02/09/2022	26/03/2023	55
37MCQ	02/09/2022	26/03/2023	56
37MCR	02/09/2022	26/03/2023	55
37MCS	02/09/2022	19/03/2023	54
37MDP	02/09/2022	11/03/2023	28
37MDQ	02/09/2022	26/03/2023	31
37MDR	02/09/2022	26/03/2023	28
37MEQ	02/09/2022	26/03/2023	36

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 in 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.

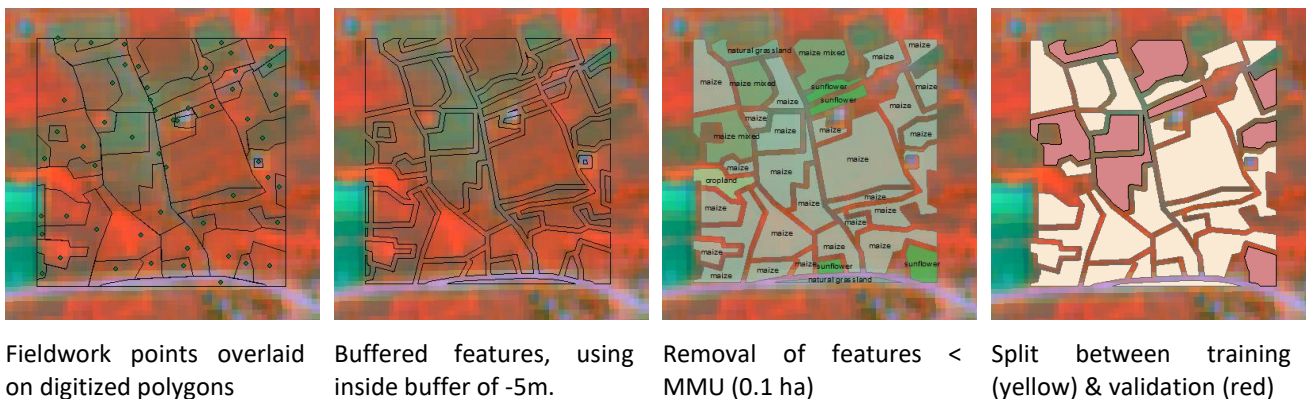


Figure 3. Preparation of fieldwork data for training and validation.

Resulting from all the described processing steps, 2,547 polygons are available for the classification process. 27% are used for training and 25% for validation. In total 51 individual classes are distinguished, of which 42 individual crop types. The figures below show a few examples from the fieldwork campaign.



Figure 4. Field with sisal & maize in mixed cropping



Figure 5. Field at an early stage of growth



Figure 6. Bare field prepared for cultivation but awaiting (belated) rainfall

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

There were no deviations from what was proposed in the feasibility study and what was done for 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). 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 CNES⁴). 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 36MXA, with a centre-date of 15-02-2023. For this synthesis, the algorithm considers all images +/- 45 days from 15-02-2023 and takes the cloud-free pixel closest to the centre date.

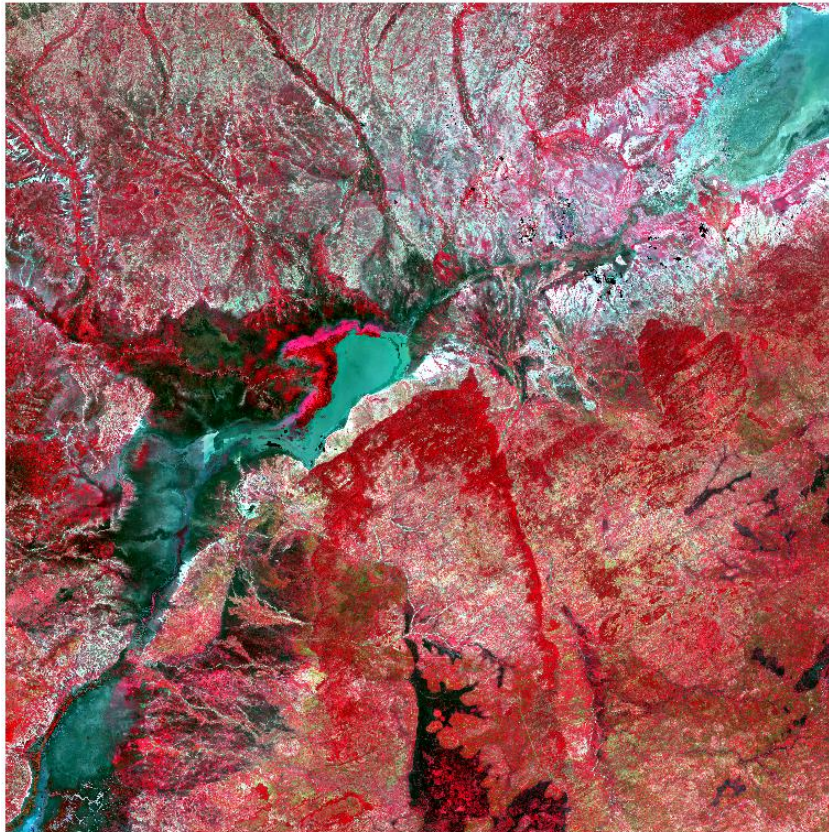


Figure 7: Sentinel-2 monthly synthesis composite (infrared false color composite), 15/02/2023, tile 36MXA.

Based on these monthly synthesis, four spectral indices are computed: the Weighted Difference Vegetation Index (WDVI⁵), 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 Tanzania when reviewing the size of the agricultural fields.

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

There were no deviations from what was proposed in the feasibility study and what was done for the in-season mapping.

⁴ <https://doi.org/10.5281/zenodo.1401360>

⁵ <https://www.sciencedirect.com/science/article/abs/pii/S092427169190005G>

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 Tanzanian short rains season mapping 2022-2023. 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 51 classes (of which 42 crop types). The Figure 8 shows the result of the raw classification output, before post-processing.

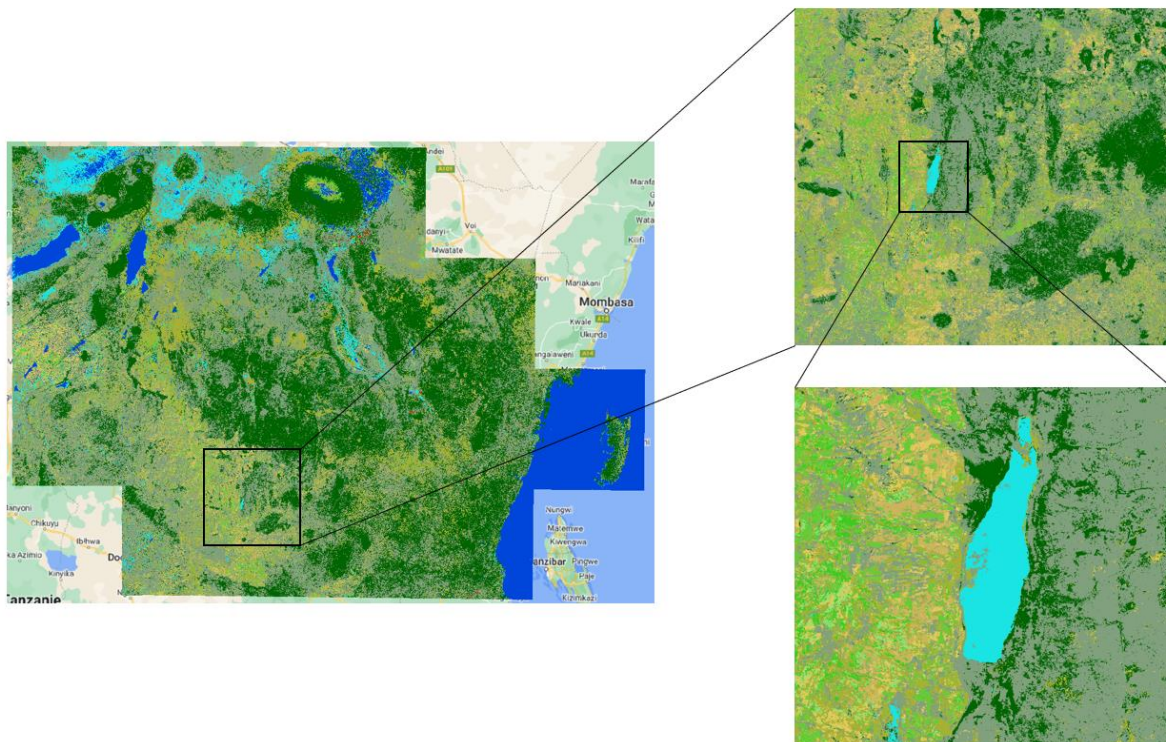


Figure 8. Raw classification output end-of-season crop type map Tanzania

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:

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, 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 51 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 (Tanzania_CropType_EndOfSeason_ShortRains_2022-2023.tif & Tanzania_CropMask_EndOfSeason_ShortRains_2022-2023.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	sunflower	Including mixed cropping with sunflower as dominant crop
11	beans	Including mixed cropping with beans as dominant crop
13	sisal	Including mixed cropping with sisal as dominant crop
14	cassava	Including mixed cropping with cassava as dominant crop
19	potatoes	Including mixed cropping with potatoes as dominant crop
6	rice	
12	peas	Including cow peas, pigeon peas, chickpeas and green peas
9	other crops	all other monoculture crops and mixed cropping as well as field preparations
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 crop 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 of course preserved in the final map. 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 36 North.

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

There were no deviations from what was proposed in the feasibility study and what was 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 for the short rains season.

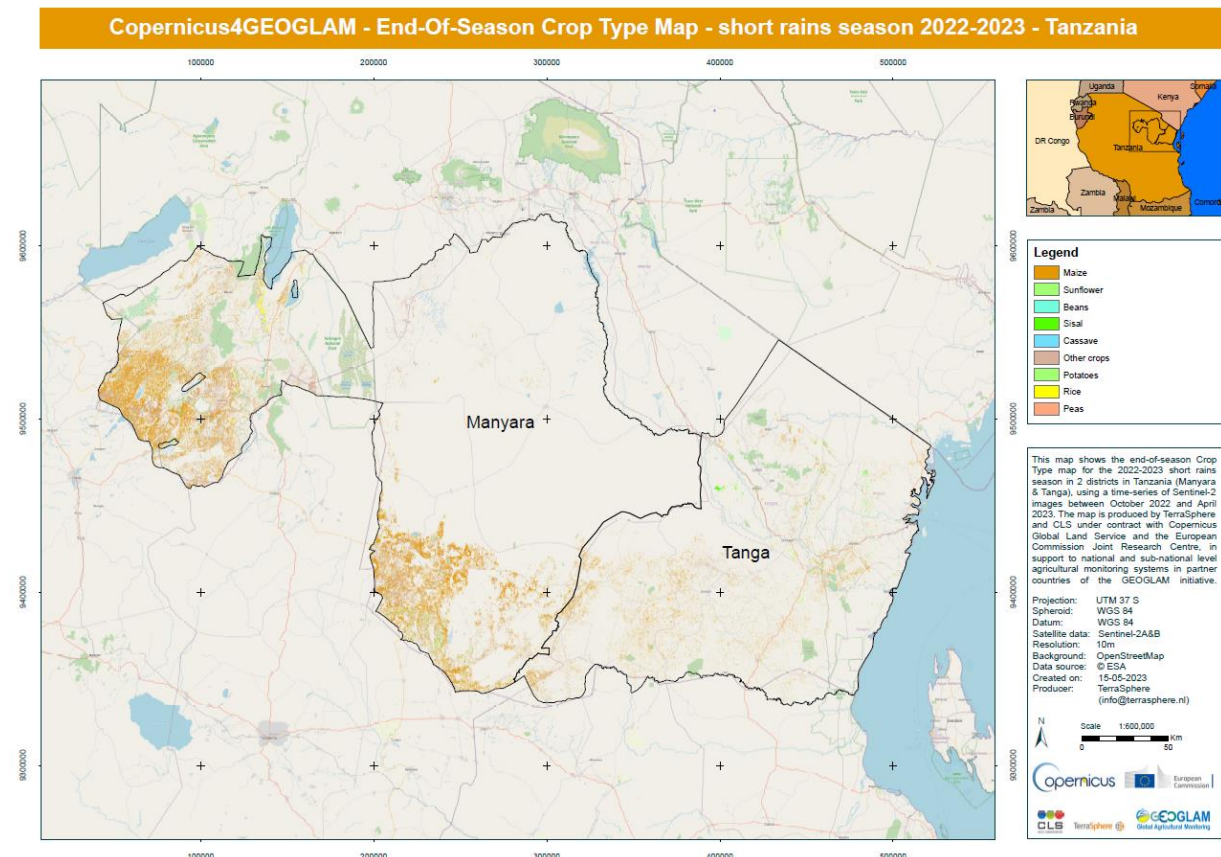


Figure 9. End-of-season Crop Mask for the short rains season 2022-2023 in Tanzania

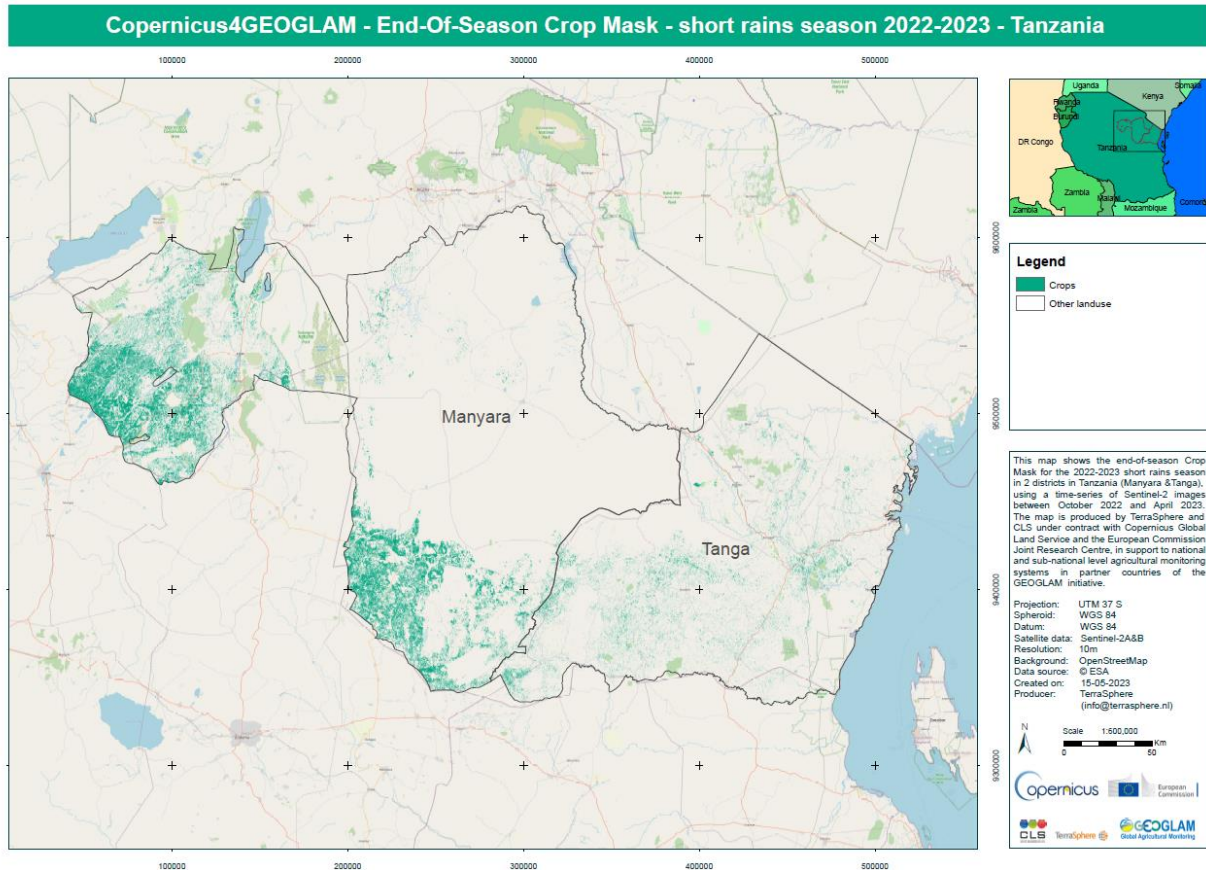


Figure 10. End-of-season Crop Type map for the short rains season 2022-2023 in Tanzania

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

There were no deviations from what was proposed in the feasibility study and what was 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 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.

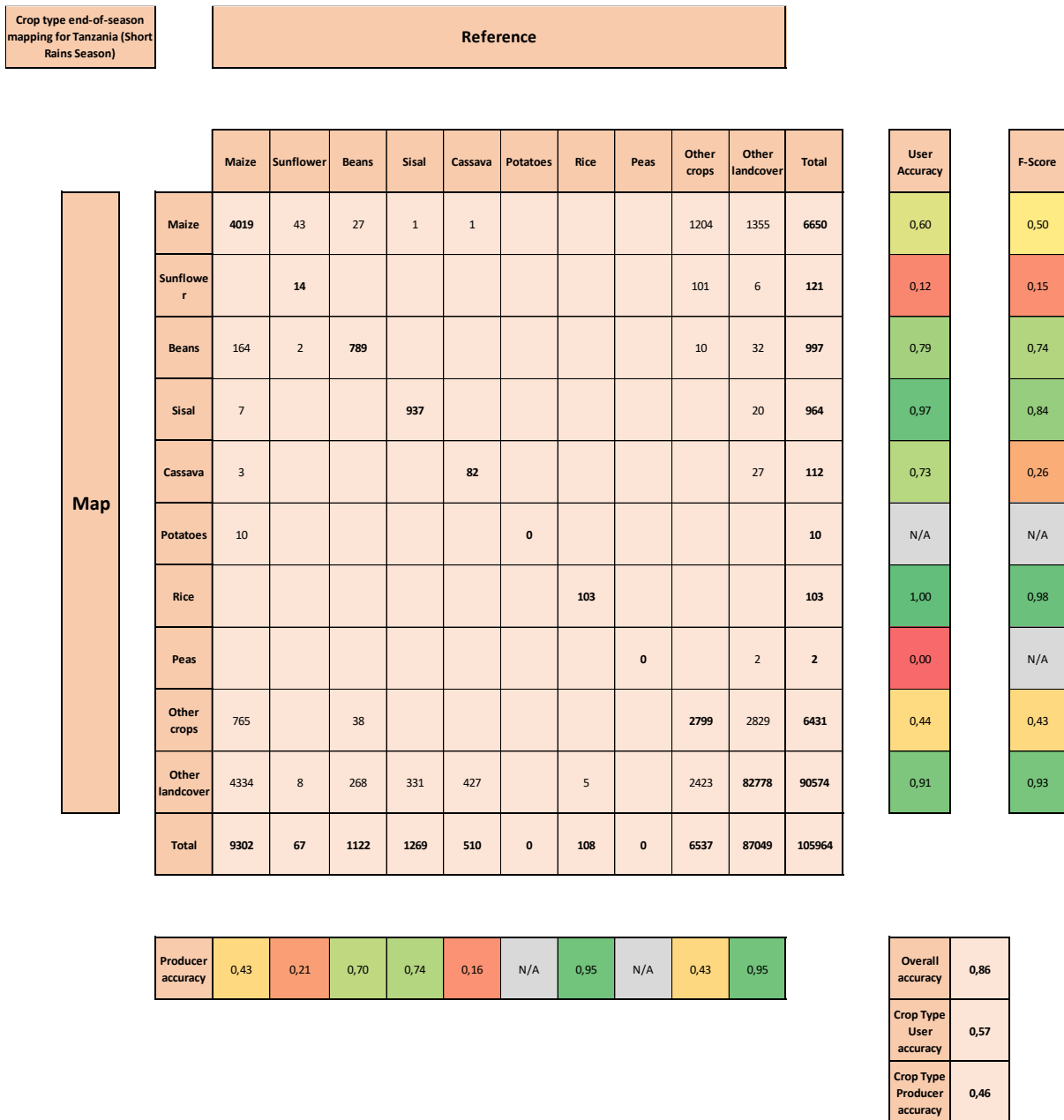


Figure 11. Confusion matrix for end-of-season Crop Type map of the short rains season 2022-2023

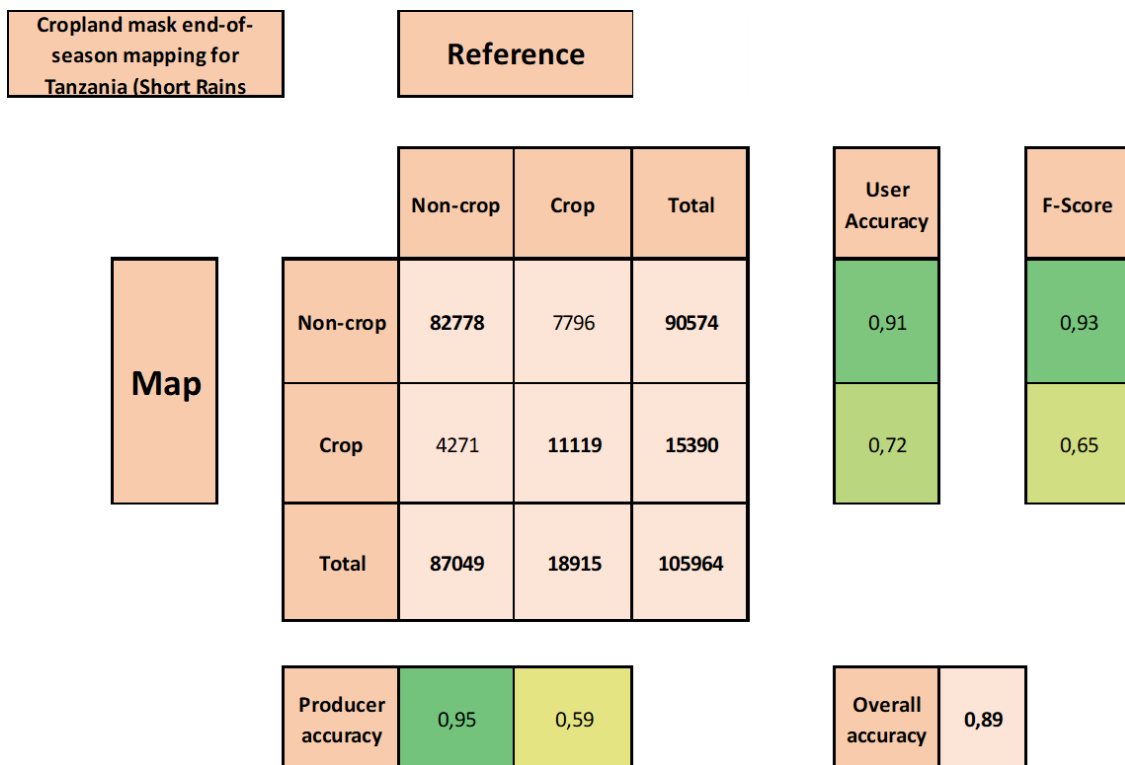


Figure 12. Confusion matrix for end-of-season Crop Mask of the short rains season 2022-2023

Figure 11 and Figure 12 show that the overall accuracy for the Crop Type map and Crop Mask is respectively reaching 86% and 89%, 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 satisfying results for the user accuracy of the crop class (72%) with 28% of commission errors. Nevertheless, the producer accuracy is low (59%) resulting in large omission errors (41%). The results obtained for end-of-season crop mask are slightly better than for in-season mapping, with a better balance between user and producer accuracy.

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”, Cassava”, “Beans”, “Sisal”, “Rice” and “Maize” classes are respectively only 40%, 21%, 27%, 3% and 0%. Overall, the “Rice”, “Sisal” and “Beans” classes tend to show satisfying results with F1-Score between 0.5 and 1. For “Sunflower” and “Cassava” classes, results are very low with F1-Score below 0.26. Overall, the results obtained for the end-of-season crop type map are lower than those obtained for the in-season map. In detail, the user accuracies are better for the in-season mapping but the producer accuracies are better for the end-of-season mapping (better balance between user and producer accuracy).

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

There were no deviations from what was proposed in the feasibility study and what was 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 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.

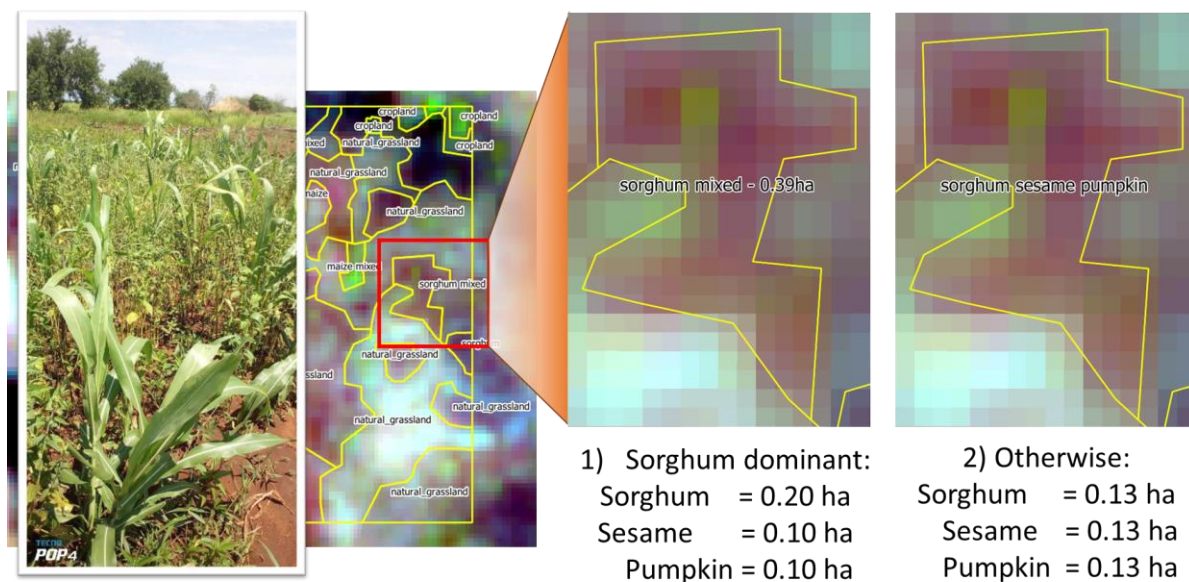


Figure 13: Mixed cropping fields and crop area estimates (non-dominant crop study case)

(2) Crop area estimates can be derived directly from the 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. It is interesting to notice the very good relative efficiencies for some crop types with figures greater than 2. For example for Wheat and Sugarcane, the same reduction in variance would have been achieved by increasing the size of the field survey sample by 6 or 19.

Table 4: Area estimates for the end-of-season mapping of the short rains season 2022-2023 in Tanzania

AOI Area (ha)		7 415 829,27										
			Maize	Beans	Green grams	Potatoes	Rice	Wheat	Sugarcane	Peas	Other crops	Other landcover
Direct Expansion	Estimate of proportion		0,08	0,01	0,02	0,01	0,01	0,00	0,00	0,01	0,06	0,81
	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	0,01	0,00	0,00	0,00	0,00	0,01	0,02
	95% Confidence Interval		0,02	0,01	0,01	0,01	0,00	0,00	0,00	0,00	0,02	0,04
	Estimate of the class area		560 978,97	64 082,96	151 080,81	80 850,61	46 956,94	2 080,80	14 912,29	42 221,15	455 126,63	5 997 538,11
	Variance		4 698 454 874,23	394 105 073,78	1 323 367 673,22	1 666 880 175,70	230 282 094,91	1 066 932,59	196 000 320,56	118 172 054,77	5 233 011 814,31	23 709 247 644,31
	Standard Error		68 545,28	19 852,08	36 378,12	40 827,44	15 175,05	1 032,92	14 000,01	10 870,70	72 339,56	153 978,08
	95% Confidence Interval		134 348,74	38 910,08	71 301,12	80 021,79	29 743,09	2 024,53	27 440,02	21 306,57	141 785,54	301 797,03
Pixel count	Map (ha)		323 009,45	8 524,55	9 308,34	6 756,67	481,35	1,62	3 749,51	127,11	163 372,33	6 900 498,34
	Map (%)		0,04	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,02	0,93
Regression Estimator	Regression estimate		0,05	0,01	0,01	0,00	0,00	0,00	0,00	0,00	0,04	0,89
	Variance		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
	95% Confidence Interval		0,01	0,00	0,00	0,00	0,00	0,00	-	0,00	0,00	0,00
	Regression estimate of the class area		398 968,62	56 202,20	92 310,82	11 775,36	32 418,21	748,02	2 339,10	31 297,49	273 535,56	6 600 251,58
	Variance		1 474 376 420,41	314 521 911,80	456 143 067,96	77 971 730,11	120 079 242,78	180 202,72	10 271 139,12	88 225 300,66	1 487 364 187,76	4 648 846 793,49
	Standard Error		38 397,61	17 734,77	21 357,51	8 830,16	10 958,07	424,50	3 204,86	9 392,83	38 566,36	68 182,45
	95% Confidence Interval		75 259,31	34 760,14	41 860,71	17 307,11	21 477,81	832,03	6 281,53	18 409,95	75 590,07	133 637,61
Efficiency	Regression Estimator		3,19	1,25	2,90	21,38	1,92	5,92	19,08	1,34	3,52	5,10

4 Conclusions

The overall accuracy for the end-of-season Crop Type map for the short rains is 86% and the end-of-season Crop Mask 89%, which is better than what was mentioned in the feasibility study (both 65%) but slightly lower than the results obtained during the in-season mapping but show a better balance between user and producer accuracy. Some classes such as “Rice”, “Sisal”, “Beans” and “Maize” show satisfying and encouraging results (F1-Score greater than 0.5) but for some individual crops (e.g. “Sunflower” and “Cassava”), lower accuracies are reported. Good relative efficiencies of 2 to 21 have been achieved for the main crops.