D4.1: Map of agricultural land of continental Africa

Project No. 869200

Version 1.0

30 November 2020

Deliverable: Soils4Africa_D4.1_v01

This project has received funding from the European Union Horizon 2020 Research and Innovation Programme under Grant Agreement 862900

Contact details

Director of Coordinating Institute – ISRIC: Rik van den Bosch

Project Coordinator: Mary Steverink-Mosugu

Address: Droevendaalsesteeg 3, 6708 PB Wageningen (Building 101), The Netherlands

Postal: PO Box 353, 6700 AJ Wageningen, The Netherlands

Phone: +31 317 48 7634

Email: mary.steverink-mosugu@isric.org

Project details

Project number	862900
Project acronym	Soils4Africa
Project name	Soil Information System for Africa
Starting date	01/06/2020
Duration In months	48
Call (part) identifier	H2020-SFS-2019-2
Topic	SFS-35-2019-2020 Sustainable Intensification in Africa

Document details

Work Package	WP4: Field Campaign and Capacity Building
Deliverable number	D4.1: Map of agricultural land in Africa
Version	1
Filename	Soils4Africa_D4-1_v01.docx
Type of deliverable	Report
Dissemination level	Public
Lead partners	IITA, RCMRD
Contributing partners	ISRIC, IRA, BUNASOLS, IFA, KALRO, SU
Author	E. Jeroen Huising (IITA); Eunice Wangui Mwangi (RCMRD); Julius Buyengo (RCMRD)
Contributors	Bas Kempen, Mary Steverink-Mosugu, Emily Toner (ISRIC)
Due date	30 November 2020
Submission date	30 November 2020

European Commission | Horizon 2020 European Union funding for Research & Innovation

This project has received funding from the European Union Horizon 2020 Research and Innovation Programme under Grant Agreement 862900

Contents

This project has received funding from the European Union Horizon 2020 Research and Innovation Programme under Grant Agreement 862900

1. Background and scope

The Soils4Africa project aims to build a soil information system that provides information about status and prevalence of soil properties and indicators relevant for sustainable intensification of agriculture in continental Africa". This information can be used as a baseline for future monitoring of soil conditions. The project is intended to support other research activities funded by the European Union Horizon 2020 programme under the umbrella of the EU-Africa Research and Innovation Partnership on Food and Nutrition Security and Sustainable Agriculture. Therefore, agricultural land of Africa defines the scope for this project. The project will collect soil samples from agricultural land only and therefore it requires a map of the agricultural area of Africa for developing the sampling scheme. The map of agricultural land of Africa is the first technical output to be delivered by the Soils4Africa project.

Agricultural land in the context of Africa has a different connotation than agricultural land has in the European context. We adopt the definition of agricultural area as provided by the FAO[1]:

Agricultural area includes arable land, permanent crops and permanent pastures.

- Arable land: land under temporary crops (double-cropped areas are counted only once), temporary meadows for mowing or pasture, land under market and kitchen gardens and land temporarily fallow (less than five years). The abandoned land resulting from shifting cultivation is not included in this category. Data for arable land are not meant to indicate the amount of land that is potentially cultivable.

- Permanent crops: land cultivated with crops that occupy the land for long periods and need not be replanted after each harvest, such as cocoa, coffee and rubber; this category includes land under flowering shrubs, fruit trees, nut trees and vines, but excludes land under trees grown for wood or timber.

- Permanent pastures: land used permanently (five years or more) for herbaceous forage crops, either cultivated or growing wild (wild prairie or grazing land).

Following this definition, FAOSTAT estimates 'agricultural land' to cover about 11.4M km^2 or about 38% of Africa's land surface area.

In practice, it is difficult to distinguish between the different land use types or to indicate the extent, or the exact boundary, of an area under a specific land use based on the interpretation of remote sensing data only. For the purpose of this project that is maybe also not required or even relevant. For example, 'land temporary fallow' is defined as land being fallowed for five years or less. For the purpose of the Soils4Africa project which concerns soil quality and sustainability of its use, it might indeed be desirable to include land that is fallowed for more than five years. For the project, areas that are part of shifting cultivation should be included in agricultural land, and from the data available to us it is difficult to look beyond the 4-year time horizon or to determine whether the land has been abandoned or not.

Another case in point is the 'permanent pastures', which refers to land that is used permanently for herbaceous forage crops, either cultivated or wild. In Africa we have extensive areas in the semi-arid region that are used for grazing, though being it used very extensively. The question, therefore, remains to which extent these lands need to be included in the mask of agricultural

[1] http://faostat.fao.org/site/377/default.aspx#ancor

European Commission | Horizon 2020 European Union funding for Research & Innovation

This project has received funding from the European Union Horizon 2020 Research and Innovation Programme under Grant Agreement 862900

land. At the same time, parts of these areas belong to the most degraded land in Africa because of overgrazing and should therefore be included in the mask of agricultural land. We would want to establish a soil quality baseline for future monitoring, certainly in those areas as well.

The subsequent question is thus whether we can identify 'agricultural land', the way we have defined it, in a reliable manner and what the best approach is to do this. The primary data source for this mapping activity are land cover maps that are generally derived from the interpretation of satellite imagery based on the spectral reflectance features. The main challenge in using this data is that the legend only provides information on land cover and not about the land use type. For instance, natural grassland is not distinguished from grassland used for grazing. Difference in spatial patterns between cultivated land and landscape that has formed naturally helps in distinguishing different land use types and differentiating between cultivated landscapes and natural landscapes. Therefore, an additional step in the interpretation of the land cover data is needed to identify agricultural land and maybe the different types of agricultural use. And for this purpose, inference rules are defined that formalizes the rules for the interpretation of the land cover data to draw conclusions on agricultural use.

The latter is of relevance to the Soils4Africa project, in that we want to be able to assess changes in soil quality in relation to its use. However, identifying the specific land use types and defining how this information is going to be used within the soil information system is outside the scope of the current activity, which primary objective is to provide a mask of agricultural area to inform the design of the Soils4Africa sampling scheme that will be used to collect soil samples from 20,000 locations across Africa. We will collect information on land use through direct observation in the field while collecting the soil samples which will provide a valuable resource for updating and improving existing land use and land cover classifications and maps.

This report describes the methodology of generating a map of the agricultural area of continental Africa and the product itself. The data is made available on the project's website. The data is provided in tiles which represent areas of 20 by 20 arc degrees. It is a dynamic repository in which subsequent tiles will be uploaded and updated when improved versions become available.

2. Methodology description

2.1. General approach

The strategy adopted by the project is to make use of existing land use and land cover data, rather to engage in an independent interpretation of satellite imagery to identify and map agricultural land. Reviewing the data and information available, we concluded that there is no justification for embarking on such an activity that would require considerable time and resources with only limited gains in the quality of the output (the map of agricultural land) given that only information on the extent of 'agricultural land' is required.

We reviewed the various data sets available (see section 2.2) and selected the one that we considered most suitable for our purpose. Subsequently, inference rules were drafted. An inference rule (or transformation rule) is a logical form consisting of a function that takes the premises, analyses the syntax and returns a conclusion (or conclusions). We adopt a broad definition of agricultural land such that the inference rules will result in an ample representation of the agricultural area, such that the errors of omission (agricultural land that is not classified as such) are limited. The implication is that areas that do not belong to the agricultural land might be included in the mask. However, because we will need to verify each proposed sampling location for its suitability before it is finally selected, this error of commission is not considered to have serious consequences.

The inference rules are drafted based on the evaluation of the land cover classification for locations with known land use characteristics. The tentative rules are applied to a subset of the data (a particular area) and the results evaluated in order to amend the rules to improve the accuracy of the classification. For the evaluation and finetuning of the rules we looked at Nigeria and the West Africa region and the Eastern Africa region comprising parts of the eastern DRC, Uganda, Kenya, Sudan and Ethiopia, because these are the areas that the team working on these rules is most familiar with. We assume the rules equally apply to the Southern, Central Africa regions, but this needs to be verified as yet. We will validate the results for the various regions and update the map as we go along. We looked specifically at areas of temporary cropping (including areas that are partially cropped), the areas with plantation crops (tree plantations, tea plantations and sugarcane plantations, for example), and areas used predominantly for grazing.

During the testing phase, the rules are applied manually in ArcGIS, using the raster calculation tool. Once the rules are final, these are implemented in Python. ArcGIS uses Python as the scripting language for the processing of geodata and offers a 'window' for entering the Python commands and running the scripts. This allows for the automated processing of the data. The data is processed in tiles, which represent areas of 20 by 20 arc degrees, and which is also the format in which the data is downloaded from the Copernicus Global Land Cover site.

2.2. Input data (Copernicus Global Land Cover data)

There are several data sets that provide information on land cover of Africa:

- Land Use and Land Cover map of Africa, 1km x 1km (30 arc seconds) resolution, based on data from 2015; data source: USGS; available through the FAO GeoNetwork portal (http://www.fao.org/geonetwork/srv/en/metadata.show?id=12755&currTab=simple)
- ESA/ESRIN land cover map of Africa 2016, 20m resolution (prototype high resolution map, based on one year of Sentinel 2A observations from December 2015 to December 2016): http://2016africalandcover20m.esrin.esa.int
- QED Cropland probability map, 1km x 1km resolution and based on 1 million GeoSurvey points of observation: (https://maps.qed.ai/map/geosurvey_h2o_crp_predictions#lat=1.56012&lng=16.75000&zoom=3.0&layers=geosurvey_h2o_crp_predictions[12
- USGS-NASA: Land Processes Distributed Active Archive Centre (LP DAAC), one of the data centres at NASA EOSDIS and part of the EROS data centre; Cropland extent 2015, 30m resolution; part of the Global Food Security-support Analysis Data (GFSAD) program: (https://lpdaac.usgs.gov/products/gfsad30afcev001/[10])
- QED cropland probability map of Nigeria, resolution not known, based on 10,000 GeoSurvey points. Data might be available for a few individual countries
- ESA, a 10m resolution worldwide landcover mapping, based on the joint use of Sentinel-1 and Sentinel-2 data. Programme has been initiated but it is not sure when the first products will be ready: https://esaworldcover.org/en

This project has received funding from the European Union Horizon 2020 Research and Innovation Programme under Grant Agreement 862900

- Copernicus Global Land Service – Land Monitoring Services: Global Land Cover 100m resolution, updates for 2019 available: https://lcviewer.vito.be/2015; https://land.copernicus.eu/global/products/lc%5B9%5D.

Furthermore, there are a number of ongoing initiatives with respect to land cover mapping that might be relevant for the purpose of mapping agricultural land or maybe for future collaboration and development of the soil information system.

- Digital Earth Africa (DE Africa), which is an operational platform intended to use available EO data and translate this in decision-ready information for Africa. This is part of Geoscience Australia, an initiative by the Australian Government: https://www.digitalearthafrica.org/

- Sentinel-2 for Agriculture, an ESA project to develop the SEN2-AGRI system intended to serve the exploitation of Sentinel-2 for local to national operational agriculture monitoring (near real time mapping): http://www.esasen2agri.org; provides access to EO data (e.g., high-resolution cloud free images from Sentinel-2 and Landsat4 imagery), but not to products like cover maps. It includes a Cropland Mask Processor for generating cropland masks at local and national scale.

- GEOGLAM (Group on Earth Observations Global Agricultural Monitoring) which is a flagship initiative under GEO (Group on Earth Observation). They are involved in global and national-regional monitoring and produce status and stock-taking reports (e.g., crop monitor): http://earthobservations.org/geoglam.php . The SOILS4AFRICA project may provide ground control data to be collected during the field campaign and contribute as such to their monitoring efforts.

After evaluating the various land cover datasets, we have selected the Copernicus Global Land Cover (CGLC) data set and maps that are produced by the Copernicus Land Monitoring Services, for the following reasons:

- It seems to be the most comprehensive product and to provide the best quality data on land cover (reliable and accurate)

- It provides the most up to date information with the most recent information provided for 2019

- It is a continuous service (the product is updated every year) rather than a one-off product for one particular year and presents therefore an opportunity for further collaboration with the Copernicus Land Monitoring Service. The annual updates also allow frequent updates of the agricultural land map. This can be easily accomplished since the mapping procedure is automated.

- It has a suitable resolution (100 m) which makes the data volume still manageable, not requiring too much computer processing capacity.

The CGLC makes use of the sensors on board the PROBA-V satellite which was designed for global vegetation monitoring. It provides dynamic Earth Observation (EO) data at 100 m resolution that is used as the primary input data and multispectral daily-synthesis surface reflectance data (300meter resolution) as the secondary source. The data is cleaned and processed to produce a 5-daily 100m time series with global coverage. Hence, the temporal resolution is very high as well, which is effectively made use of for the classification of land cover.

The CGLC provides land cover classification map provides 23 discrete land cover classes, in which Forest is divided into 'Open Forest' and 'Closed Forest' at the second level of the classification hierarchy and each of these two forest classes further subdivided into 6 classes according to whether it is evergreen or deciduous, broad leave or needle leave, mixed or unknown. Apart from information on the discrete land cover classes, the CGLC provides information on the fractional cover for each of the primary classes. The fractional cover is obtained through a rather complicated process in which the training data is generated using 10m resolution satellite imagery from which the composition of the 100 by 100m cells of the CGLC is determined (Buchhorn et al., 2020).

The following primary classes are identified:

- Forest
- Shrubland
- Herbaceous vegetation
- Moss & Lichen
- Bare / Spars vegetation
- Cropland
- Built-up
- Snow & Ice
- Permanent water bodies

For further information on the "moderate dynamic land cover – 100m", the algorithms used for classification and regression analysis and product description, please see Buchhorn et al. (2020).

In addition to the CGLC map, we used the Agroecological Zones (AEZ) to define the inference rules for mapping agricultural land. We use the definition by the FAO that has been adopted by IFRPI (HarvestChoice/International Food Policy Research Center (IFPRI), 2015). In Figure 1, the AEZs are displayed. To be used in the process of identifying and mapping agricultural land we need to extract the AEZ data for each tile, a geographical area of 20 by 20 arc degrees, by which the data is being processed. Further the map needs to be resampled to a 100 by 100-meter resolution to be commensurate with the resolution of the global landcover data.

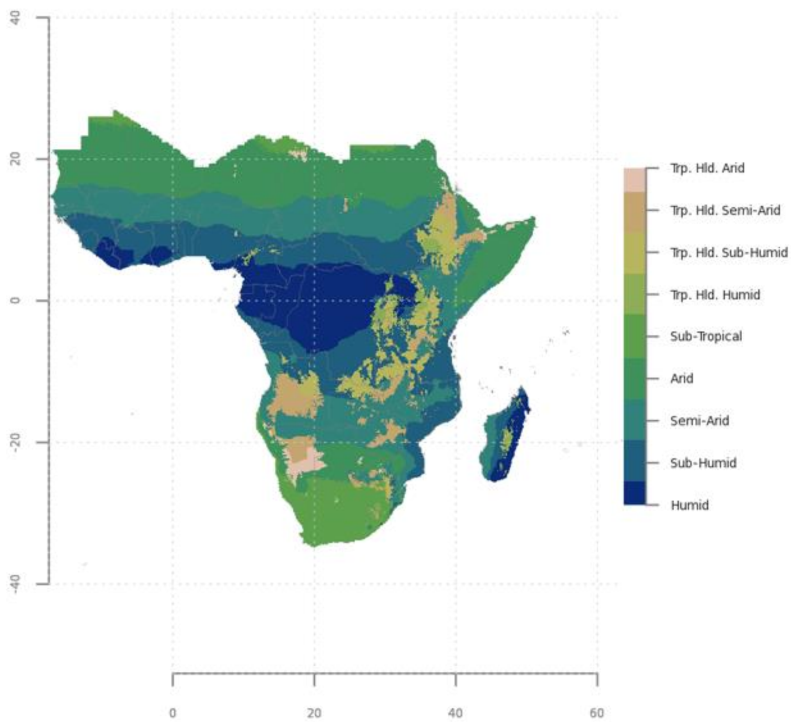

Figure 1. Agroecological zones (8-classes) for Sub-Saharan Africa (HarvestChoice/ International Food Policy Research Center (IFPRI), 2015).

European Commission | Horizon 2020 European Union funding for Research & Innovation

This project has received funding from the European Union Horizon 2020 Research and Innovation Programme under Grant Agreement 862900

2.3. Inference rules

Inference rules were drafted by evaluating the classification result of the CGLC for locations with known land use, whereby images with a 10m resolution from Google Maps were used for reference. The presence of a field pattern is the most important criterion hereby. For an area with the same land use pattern, we look at the discrete land cover class and the fractional cover of the various land cover classes for several locations within that area, such that we get insight in the ranges in fractional cover that are characteristic for the particular land use. Hence, inference rules were developed using an area-based approach rather than a point-based approach.

We use a two-pronged approach. In the first approach we took the discrete land cover class as the starting point and subsequently evaluated the land use pattern from the visual interpretation of the Google Maps imagery and subsequently we looked at the fractional cover percentages for the various land cover classes. The focus was on "Cropland", "Forest", "Shrubland" and "Herbaceous vegetation" cover classes. However, also classes like "Wetland herbaceous" were investigated whether they might represent agricultural area. The second approach was to start with assessing the land use pattern for a particular, generally larger area using Google Maps and for which the land use type is known and to subsequently assess the discrete land cover classes and the fractional cover percentages for that area. Hereby, the focus was on areas of continuous cropland, areas with partial cropland (no contiguous field pattern), different types of plantations and areas of deforestation or agricultural colonisation areas in which we find a pattern of isolated plots or parts of land that have been deforested (within a forested area).

In drafting the rules, we took account of the agroecological zone (AEZ) as well, because it provides relevant contextual information for the interpretation of the land cover classes. The land use pattern as well as the type of vegetation differs characteristically between the various AEZs. (E.g., tea plantations are not found in semi-arid regions). The AEZ is therefore used as auxiliary information to refine some of the inference rules.

The draft inference rules were subsequently implemented in ArcGIS and applied to the input maps, and the result validated again by using satellite imagery (Google maps) for ground truthing. Based on the ground truth observations the rules were adjusted. This was an iterative process through which the rules are refined. After the final round of validation, the rules were formulated as IF-THEN statements which output returns either the value "true" or "false" with respect to the statement that the area (pixel) belongs to the class "agricultural land". The following rules have been established. An explanation is provided with each rule provided.

IF SLC class = "Cropland" THEN MAL = "1" [1]

where SLC class is the land cover class in the discrete land cover classification map, and MAL stands for *Mask of Agricultural Land*. It has a value "1" if true and "0" if false. We find that if the cell is classified as "Cropland" in the discrete land cover classification map it is very likely cropland, irrespective of the fractional "Cropland" cover percentage and irrespective of the AEZ. In some cases, the fractional "Cropland" cover might be quite low. Also, if the fractional "Cropland" cover is 0% it positively is not cropland and can be excluded from agricultural land.

IF SLC class = "Closed Forest" AND FLC-FO ≥ 50% THEN MAL = "0" [2]

This project has received funding from the European Union Horizon 2020 Research and Innovation Programme under Grant Agreement 862900

where FLC-FO is the fractional forest cover.

For area classified as "Forest" cover it makes sense to distinguish between "Closed Forest" and "Open Forest" (second level in the CGLC classification hierarchy). If the discrete class is "Closed Forest" and the Fractional "Forest" cover percentages is ≥ 50% it is positively identified as forest and can, therefore, be excluded from the agricultural land mask. We observe that patches where the forest has been cleared are well captured and are recognised as agricultural land applying the rules presented here. We see larger areas in the 'derived savanna' zone area that classified as forest, but that are not forest in actual sense. We might observe the same in the 'humid forest' AEZ and this rule helps to distinguish between the actual forest and areas with a high tree cover but that are not actual forested area.

IF SLC class = "Closed forest" OR SLC class = "Open forest" AND FCL-FO ≥ 30% AND FCL-SL ≥ 30% AND FCL-CL < 15% THEN MAL = "1" [3]

where FCL-SL is fractional "Shrubland" cover percentage and FCL-CL is the fraction "Cropland" cover.

We find tree plantations (e.g., teak plantation) of which the grid cells are being classified as either "Closed forest" or "Open forest", but that is characterised by a considerable "Shrubland" fractional cover percentage and that distinguishes it from forest area. It has to show a "Cropland" fractional cover of less than 15% to distinguish it from land under temporary crops. Alternatively, we can assign a separate value for "Plantation tree", if it is important for the sampling design as well as for the design of the SIS to identify the specific land use class. But in that case, we will need an independent validation of the classification result to establish which land use and vegetation types this class actually represents.

IF SLC class = "Closed forest" OR SLC class = "Open forest" AND FCL-FO ≤ 50% AND FCL-CL ≥ 15% THEN MAL = "1" [4]

This rule applies to the 'humid forest' and the 'derived savanna' agroecological zone especially. We find larger contiguous areas that are classified either as 'Open Forest' or 'Closed Forest' (or alternately as 'Closed' or 'Open Forest' in some cases) that actually belong to agricultural land, which in part is cropped and in other parts is not. These areas we find as large 'buffer' zones (concentric circles) around the major settlements and larger cities. Alternatively, you could say these areas are classified as "Open Forest" in which you still find large expansions classified as 'Closed Forest'. We find crop land, but not as contiguous fields. The field pattern generally shows a patchy distribution, rather than being dispersed. The land that is not cropped is left fallow or not suited for annual cropping and used for grazing purposes or other. However, the land has been deforested (undergone conversion of land use) at some time in the past and the whole zone should be designated as 'Agricultural land' and is considered relevant for monitoring soil quality.

We do find few patches classified as 'Shrubland' and 'Herbaceous vegetation', which should be included in the masks of agricultural land, especially if associated with areas classified as 'Cropland'. These areas may indeed relate to shrubland or 'wooded area' that we may find along streams and drainages. In so far this "Shrubland" belongs to agricultural area we assume these areas are accurately captured by the rule [6] that considers the area classified as "Shrubland".

European Commission | Horizon 2020 European Union funding for Research & Innovation

This project has received funding from the European Union Horizon 2020 Research and Innovation Programme under Grant Agreement 862900

We do find grid cells that have a fractional 'Cropland' cover of less than 15% and that still belong to agricultural land. But if we use a threshold value of less than 15% the rule would possibly conflict with rule 3 and it would not be of any consequence because the area would still be classified as agricultural land because of rule 3.

We can classify this as 'partial cropland area' and assign a different value (e.g., 1.2), if it is important to identify the different land use types.

IF SLC class = "Wetland herbaceous" THEN MAL = "0" [5]

Area classified as "Wetland herbaceous" in the discrete land cover classification is indeed natural wetland area, irrespective of cells having some percentage of cropland cover, or of 'herbaceous vegetation'. We investigated this land cover class because "Herbaceous vegetation' may refer to grasslands that are used for grazing purposes. However, that does not seem to apply in this case and should be excluded from 'agricultural area', even though it may quality as such based on the fraction cover percentages.

IF SLC class = "Shrubland" AND FCL-CL ≥ 25% AND AEZ = "Semi-humid" THEN MAL = "1" [6]

where AEZ stands for the agroecological zone definition according to IFPRI.

In the derived savanna AEZ and the Southern Guinea savanna zone (the sub-humid zone) we find agricultural land that is classified as "Shrubland" in the discrete land cover classification. The fractional "Cropland" cover ranges between 25% and 40%. This corresponds to agricultural land with partial cropland (temporary crops) with patchy distribution of the agricultural fields. It is a mosaic of patches of cultivated and non-cultivated land with no clear field pattern in the areas that are not cultivated. Actual cropland can easily constitute up to 50% of the surface area. This rule applies to the Southern Guinea savanna and derived savanna zone and not for the Northern Guinea savanna or the Sudan savanna (the semi-arid zone). The AEZ is, therefore, included as an additional criterion in the inference rule, which classifies the former as 'tropical – semi-humid'... For the semi-arid area this rule does not apply, even though the conditions in the rule may be fulfilled. We assume this rule applies equally for the West Africa, East Africa and Southern Africa region.

We can designate these areas as "partial crop land area" for the Guinea savanna" and assign a separate code (e.g., 1.3) to distinguish it from the partial cropland for the derived savanna that result from applying rule 4

IF SLC class = "Shrubland" AND AEZ = "Semi-arid" THEN MAL = "0" [7]

In the tropical semi-arid agroecological zone, areas that are classified as "Shrubland" should not be classified as agricultural land. We do find factional "Cropland" cover percentages of more than 25%, but we cannot discern any field pattern in these areas to support such classification. The area is possibly used for grazing, but this will be very extensively used. This rule is included to

This project has received funding from the European Union Horizon 2020 Research and Innovation Programme under Grant Agreement 862900

differentiate this area from the one defined by rule 6. That is if rule 6 is applied without the additional condition of the AEZ, then this rule should be applied to exclude those area satisfying the land cover conditions within the semi-arid areas.

Alternatively, we might assign the class name "Dryland shrub vegetation" and assign a separate class code (e.g., 3), if possibly at later instance we might decide to include this area in the mask of agricultural land as wild prairie or grazing land for the design of the sampling scheme.

IF SLC class = "Herbaceous vegetation" THEN MAL = "0" [8]

We find areas classified as "Herbaceous vegetation" in the dryland zone mainly (Sudan savanna mainly). We find signs that people have settled in this area. Some settlement patterns are visible, and we may observe some vague patterns that look like agricultural fields. In case there are fields that are cultivated the discrete landcover classification seems to identify those accurately. The area will be used for grazing, but this will be very extensively managed. We therefor want to exclude these areas from agricultural land. We observe fractional "Cropland" cover percentages up to 40% but this does not adequately reflect the actual cropland percentage.

We could assign a separate class "Dryland herbaceous vegetation" to the area that fulfils the conditions and give it a class value of "4", for example, for possible future consideration to be included in the mask for designing the sampling scheme.

Possibly there may be areas within the 'derived savanna' and 'southern Guinea savanna' that are classified as "Herbaceous vegetation" and with a fractional "Cropland" cover of up to 40% that does represent agricultural area. It seems to hardly occur, though. We nevertheless include such a rule to be safe.

IF SLC class = "Herbaceous vegetation" AND AEZ = "Semi-humid" AND FLC-HV ≥ 25% THEN MAL = "1" [9]

where FLC-HV is the fractional "Herbaceous vegetation" cover.

This rule is included to identify the true grazing land, whether cultivated or wild that is managed to some degree and that should be included in the mask of agricultural land, therefore. Alternatively, we can assign a separate class value (e.g., 1.4) and name it "Semi-humid herbaceous" for if we consider further differentiation of the agricultural land according to land use type useful for future purpose.

IF SLC class = "Cropland" AND FLC-HV ≥ 25% THEN MAL = "1" [10]

We observe that sugar cane plantations, for example, are generally classified as "Cropland", with a relatively high fractional cover of "Herbaceous vegetation". However, the FLC-HV generally remains below 40%. Tea plantations may be partially identified by this rule as well. This rule provides for a further subdivision of cropland area that is identified by rule [1]. We could classify

European Commission | Horizon 2020 European Union funding for Research & Innovation

This project has received funding from the European Union Horizon 2020 Research and Innovation Programme under Grant Agreement 862900

this a "Plantation crop - herbaceous" and assign the value 1.5 fi we want to adopt a second level in the classification of 'cropland'.

IF SLC class = "Forest" AND FLC-HV ≥ 25% THEN MAL = "1" [11]

Tea plantations are generally classified as "Forest" in the describe land cover classification (mainly as "Closed forest", though "Open forest" occurs as well), with a considerable percentage of the fractional cover of "Herbaceous vegetation". We see that FCL-CL may vary between 10% and 30%. That means there is some overlap with the area classified by rule [4] but it is an addition to rule [4] as well in that it identifies area belonging to "Plantation crop – woody plants" to be included in the second level of the classification hierarchy of agricultural land.

2.4 Implementation of the inference rules and processing

One way of implementing the rules in ArcGIS is by using the raster calculator tool. For this purpose, each condition is evaluated separately resulting in a binary map layer, with "1" indicating the condition is fulfilled ("TRUE") and "0" indicating the condition is not met ("FALSE"). The various data layers (binary map layers) are subsequently combined using Boolean operators in which way the complete rule is evaluated (returning a value "TRUE" or "FALSE"). This is done for each rule separately. Subsequently the results for all rules are combined using the Boolean "OR" operator. This returns a "TRUE" value for each grid cell for which at least one of the input maps has been evaluated "TRUE" and returns a "FALSE" only in those cases where all input maps (rules) are evaluated "FALSE".

The process is illustrated in Figure 2 for rules [1] and [3] and the combination of both rules. Hereby MAL1 stands for the resulting Mask of Agricultural Land for rule [1] with 'MAL3' being the same for rule [3]. MAL stand for the Mask of Agricultural Land that results in the end. 'LC100-global_v' stands for the discrete land cover classification in which the various classes have a unique identifier (number). 'cls_op_for' is the binary file which indicates the area classified as "Open forest" in the discrete land cover classification map. 'forest30g' is the binary file that result from evaluating the fractional forest cover being more than or equal to 30%, 'shrb30g' being the binary map indicating area with fractional "Shrubland" cover or more or equal to 30% and 'crop15l' being the binary map which indicates areas with fractional "Cropland" cover of less than 15%.

This project has received funding from the European Union Horizon 2020 Research and Innovation Programme under Grant Agreement 862900

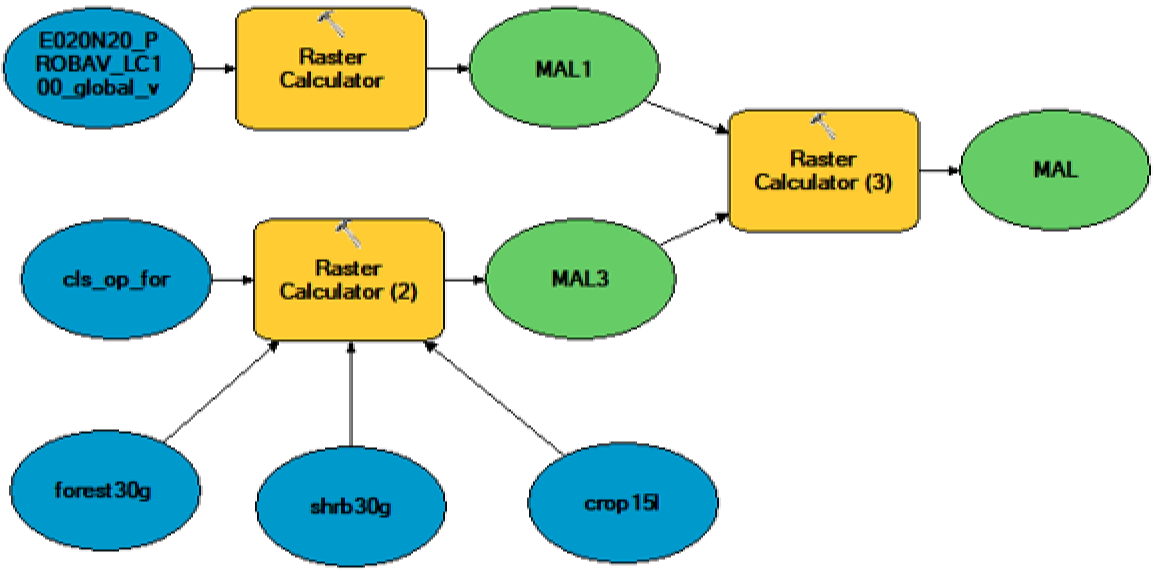

Figure 2. Schematic representation of inference rules [1] and [3].

Rules that identify areas that need to be excluded from the mask of agricultural land need to be handled differently. An example is rule [2] that identifies the 'true' forest areas. If the rule is exclusive (that is, there is no other rule that might identify an area as agricultural land that is at the same time identified as forest (or non-agricultural land) than this rule does not need to be executed. If the rule is not exclusive, then the rule needs to be implemented. Evaluating rule [2] would give a "TRUE" value ("1") for forest areas and this needs to be excluded by using a Boolean NOT statement in the final evaluation in which the outcomes of all rules are considered together. This applies for rule [7] and [8], which if not applied would result in area incorrectly classified as 'agricultural land' as well.

3 Description of the map output

3.1 Output specifications

The mask of agricultural land of continental Africa only provides information on where land is used for agricultural purposes as defined by the FAO. It is a binary map in which "0" indicates area not used for agricultural purpose and "1" indicates use for agricultural purpose, without giving further detail on the specific type of land use (though that information is existing and might be made available at later stage; see section 4 on future activities).

The data is made available in tiles of 20 by 20 arc degrees as indicated in Figure 3. These are the same tiles by which the Global Land Cover data is made available from the Copernicus Land Monitoring Service[2]. There are 13 tiles in total needed to cover the whole of the African continent, including Madagascar. The projection used is the Goode Homolosine which is an equal area pseudo-cylindrical projection. It is widely used for world maps and supported by ArcGIS[3] . The

[2] https://lcviewer.vito.be/2015

[3] https://pro.arcgis.com/en/pro-app/help/mapping/properties/goode-homolosine.htm

European Commission | Horizon 2020 European Union funding for Research & Innovation
This project has received funding from the European Union Horizon 2020 Research and Innovation Programme under Grant Agreement 862900

'equal area' feature makes that the surface area of agricultural land as indicated by this map is an estimate of the actual 'true' area of agricultural land. Furthermore, an equal area projection with a metric unit is required for the sampling design. The number of the tile is for reference purpose and is also referenced in the file name to identify the corresponding geographical area the map covers. The maps are referred to as map of agricultural land (MAL).

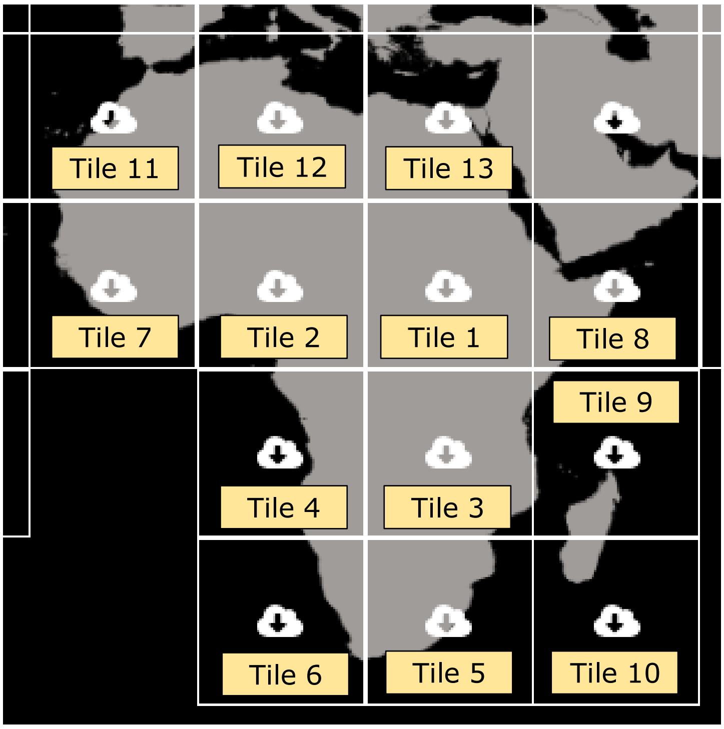

Figure 3. Copernicus Global Land Cover map tiles (20 arc degrees * 20 arc degrees). Each tile has an identification number that is also referenced in the corresponding file name.

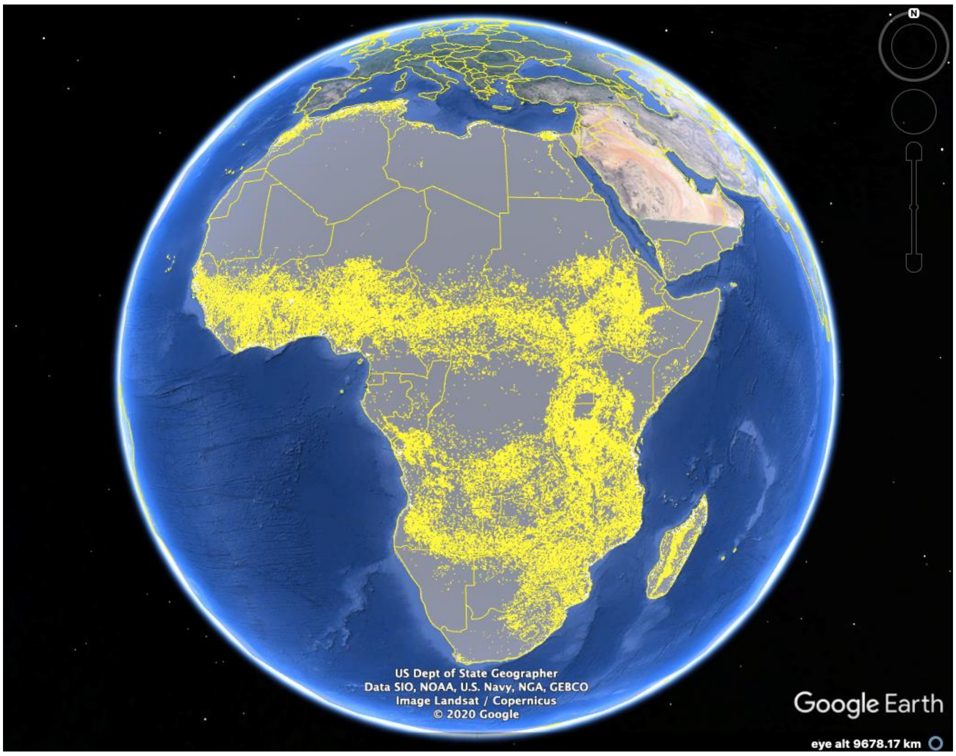

Figure 4. Mask of agricultural area for continental Africa, overlain in Google Earth on the map of Africa (©Google Earth).

European Commission
Horizon 2020
European Union funding for Research & Innovation
This project has received funding from the European Union Horizon 2020 Research and Innovation Programme under Grant Agreement 862900

The maps are made available in GeoTIFF file format. A map of agricultural area of the whole of the African continent is also made available. Figure 4 displays the agricultural area (in yellow) of continental Africa overlain in Google Earth on the map of Africa. The image resolution is 100 by 100 meters. The total surface area of agricultural land in Africa according to our calculations is 7.925M km^2, which is considerably less than the 11.4M km2 specified by the FAO. The total land area of continental Africa according to our estimation is 30.937M km2 and the agricultural area therefore covers 25.6% of the surface area of continental Africa.

A further explanation of the classification results is given in the section below in which the result of the mapping of agricultural land is evaluated.

3.2 Evaluation of the agricultural land map

Here we provide further details on the mapping of agricultural land, illustrated by the results obtained for Tile 1. Tile 1 covers the geographical area from the center of the Democratic Republic of Congo in the southwest to the Red Sea off the coast of Eritrea in the northeast. It spans most of the agroecological zones of Africa from the 'tropical humid forest in DRC to the 'tropical-arid regions of Sudan and the 'Tropical Highland' areas of Ethiopia. Also, most of the land use types typical of Africa are represented in this area.

Figure 5 (left) gives the mask of agricultural land (displayed in red) overlaid on the satellite data of Google Maps. Figure 5 (right) gives satellite data only for reference purposes. The mask of agricultural land is the result of the addition of the area that results from evaluating each of the inference rules.

Figure 6 gives a detail of the mask of agricultural land and the corresponding area in Google Maps for the area around Lake Turkana and Marsabit in the northwest of Kenya. Marsabit and Lake Turkana are located in the semi-arid region. Around the settlements like Marsabit we find areas that are correctly classified as agricultural land, being land that is used for cropping.

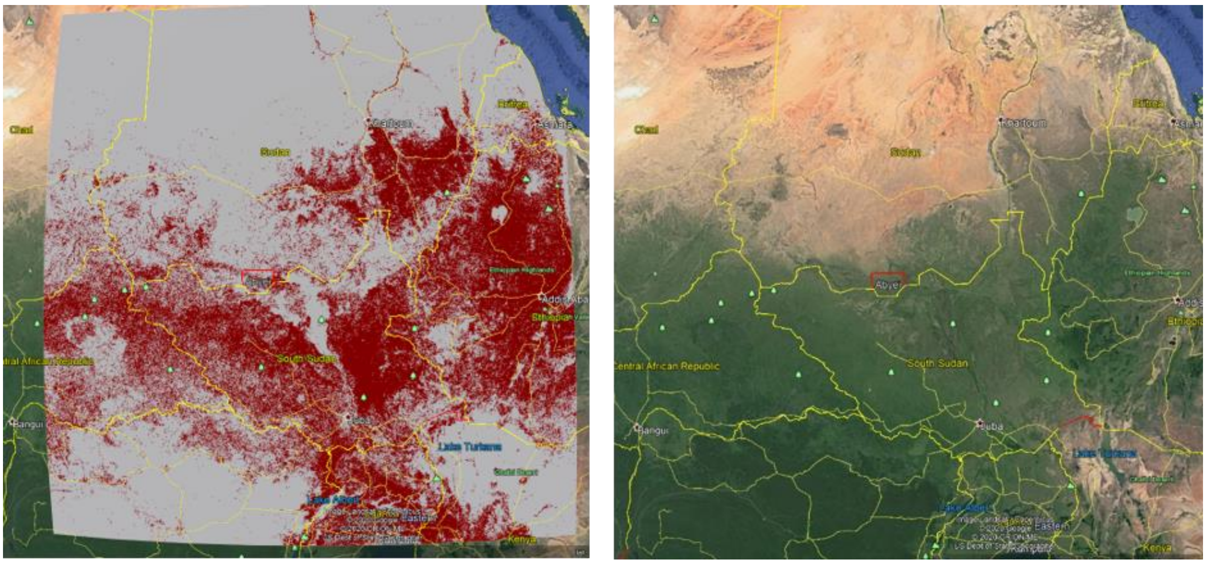

Figure 5. Map of agricultural land for Tile 1 overlain on satellite imagery of Google Maps (left) and corresponding area for refence and orientation (right).

This project has received funding from the European Union Horizon 2020 Research and Innovation Programme under Grant Agreement 862900

We also find a lot of isolated grid cells classified (correctly) as agricultural land, and the question is how the algorithms used for the sampling design will deal with these isolated cells. However, removing those isolated cells (by using a spatial filter) will reduce the surface area of agricultural land considerably. These isolated cells are retained in the mask of agricultural land and we will leave it for later to decide how we deal with these cells in the process of designing the sampling scheme.

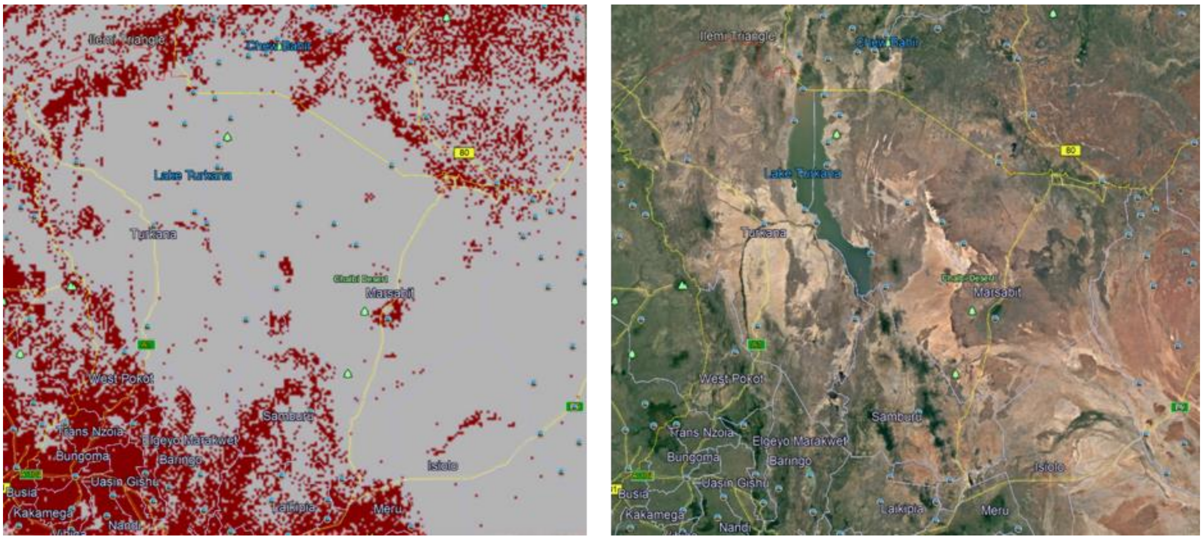

Figure 6. Detail of the map of agricultural land for the area in the northwest of Kenya, around Marsabit and Lake Turkana (left) and the corresponding satellite imagery for the same area (right).

Figure 7 shows the area of Tile 1 that is classified as "Cropland" in the CGLC, and which results from applying inference rule [1]. It identifies the areas southeast of Khartoum (includes Gezira irrigation schemes), the Ethiopian highlands and the northern part of Uganda as the main cropland areas. It is only a fraction of the total area of agricultural land for this geographical area.

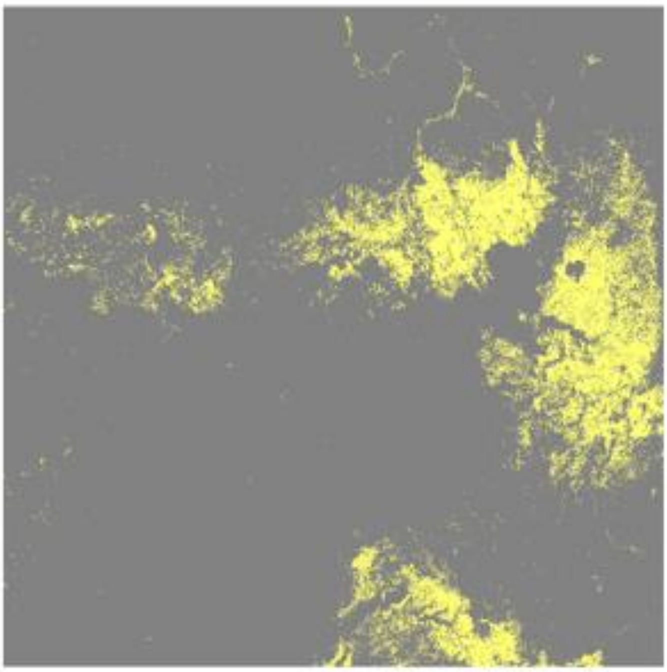

Figure 7. The area of Tile 1 classified a "Cropland", resulting of the application of inference rule [1].

This project has received funding from the European Union Horizon 2020 Research and Innovation Programme under Grant Agreement 862900

Figure 8A results from applying inference rule [2] and indicates dense and closed forest vegetation. It is excluded from the mask of agricultural land. Figure 8B result from the application of rule [3]. It is area classified as forest but with a high fractional cover classified as "Shrubland". When we find these areas at the forest margins and forest fringe, they are associated with secondary forest areas; areas that have been deforested at some time in the past or forests that have been severely logged. It may also be associated with (old) tree plantations (like teak plantations in Nigeria for example) that are characterised by a smooth canopy structure. Though it remains a matter of debate whether this should be included in the mask of agricultural land, but so far we did. It certainly signifies a very extensive use of the land. Where these areas are found in isolation (not in direct neighbourhood of forest areas) these will correspond to tree plantations indeed.

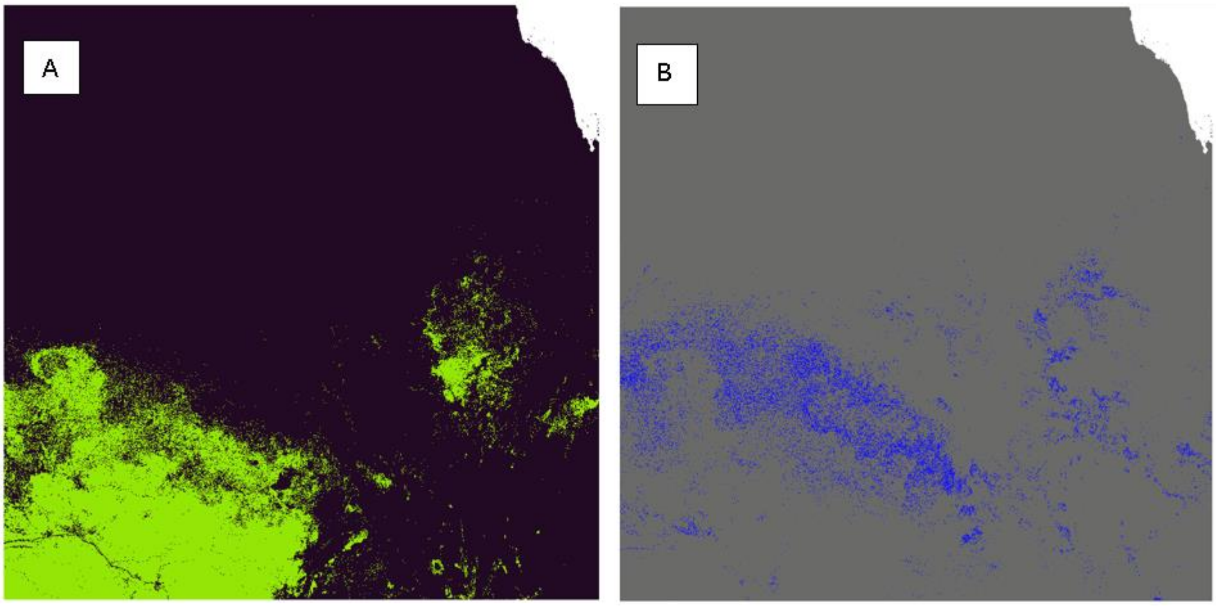

Figure 8A and B. Close dense forest area resulting from inference rule 2 (A) and area with dense woody vegetation with smooth tree canopy resulting from inference rule 3 (B).

The area in Figure 9 is classified as "Open Forest" with a relatively high fractional cover percentage of "Herbaceous vegetation" and consequently a low percentage of fractional "Cropland" cover. This area results from the evaluation of inference rule [11]. The area belongs to the sub-humid tropics but tends towards the drier conditions. We do not find actual forests in these areas and neither do we observe fields or a field pattern in most of the area. A lot of these areas will be naturally grazed, and we do find a number of large natural parks, wildlife reserves and natural reserves in this zone and these areas should be excluded from the agricultural land mask used for defining the sampling scheme. Otherwise, agricultural use in this zone will be for grazing, but it will be extensively used and seems not to be managed. Therefor the question remains whether this should be included in the mask of agricultural land. We have included it in the mask thus far. We might want to consider using a much lower sampling density for this area to reflect the limited importance and relevance of this area for agricultural production.

The areas of Figure 10A and B are complementary to each other and have a comparable land use pattern, namely that of partial cropland area. The area of Figure 10A is classified as "Open" and "Closed forest" and the area of Figure 10B as "Shrubland", but both with a significant percentage of "Cropland" fractional cover. The areas together take an intermediate position between continuous cropland area and the forest areas. The rules seem to provide a reliable result irrespective of the agroecological zone being part of the 'tropical region' or the 'tropical highland' region. These areas are considered to belong to the agricultural area without any doubt.

This project has received funding from the European Union Horizon 2020 Research and Innovation Programme under Grant Agreement 862900

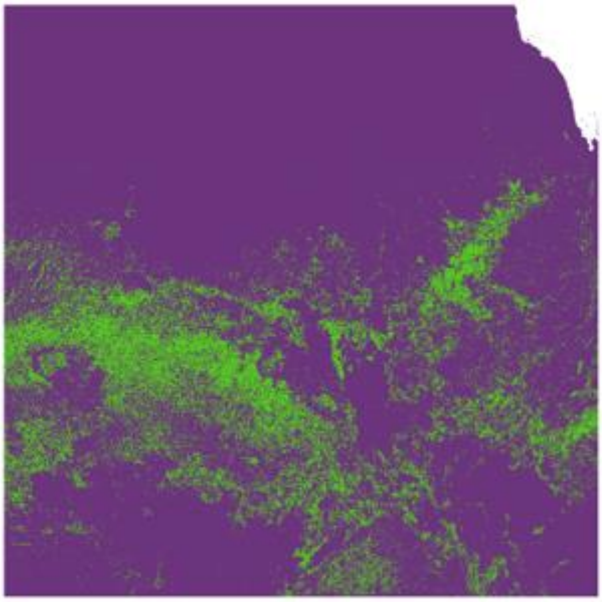

Figure 9. Area classified as “Open forest” with a high percentage of fractional cover of “Herbaceous vegetation”.

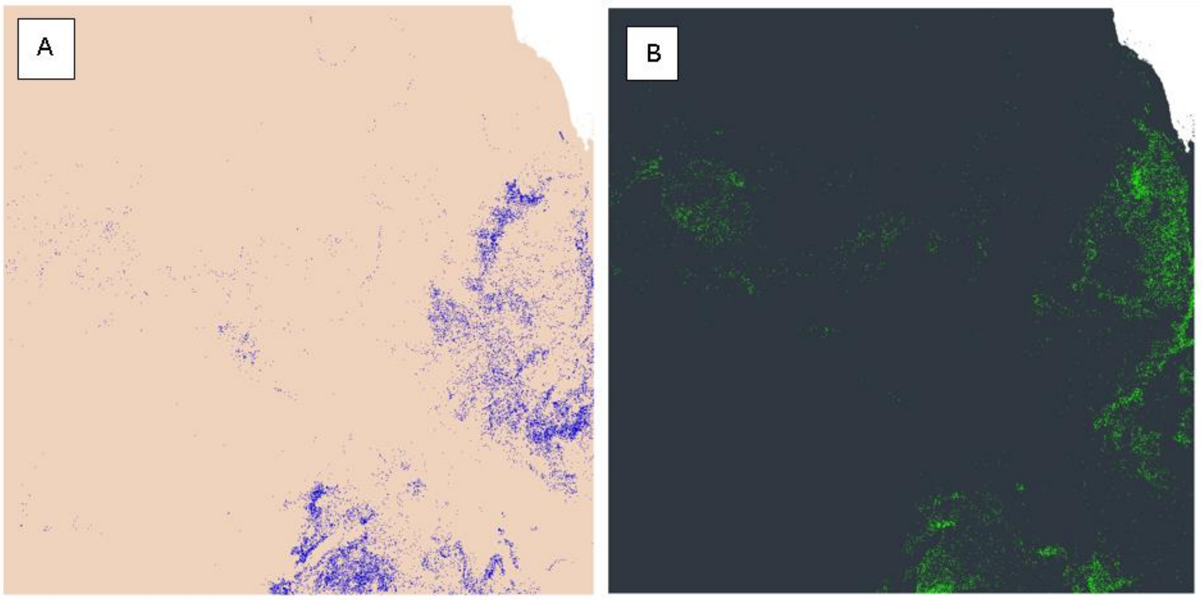

Figure 10A and B. Partial cropland area of the sub-humid (derived savanna) region resulting from the evaluation of inference rule [4] (A) and partial cropland area for the sub-humid (southern guinea savanna) region resulting from the evaluation of inference rule [6] (B).

Inference rule [9] identifies areas with herbaceous vegetation within the sub-humid region. The area indicated in Figure 11A corresponds with the extensive grazing areas in the southeast of South Sudan. We do not observe any field pattern indicating that cropping might take place. The area is used extensively but seems rightfully included in the mask of agricultural land. The area indicated in Figure 11B denotes the area classified as “Wetland herbaceous” and it corresponds very well with location of the Sudd in South Sudan and is rightfully excluded from the mask of agricultural land.

This project has received funding from the European Union Horizon 2020 Research and Innovation Programme under Grant Agreement 862900

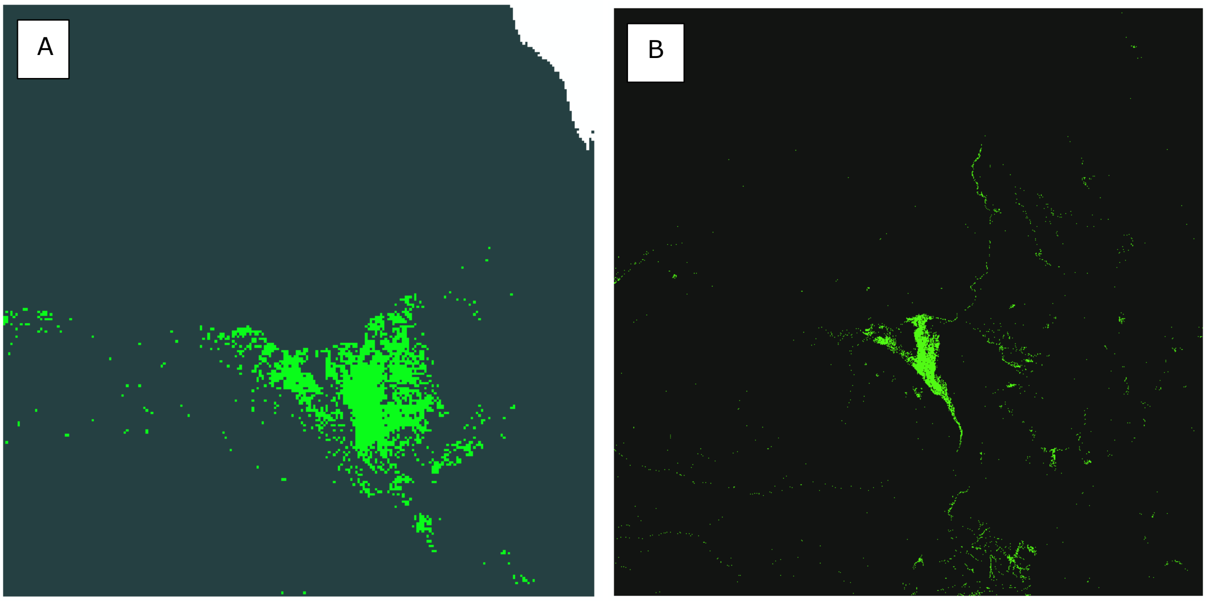

Figure 11A and B. Area with predominantly herbaceous vegetation in the sub-humid region (A) and the area classified as Wetland herbaceous (B).

5. Future work

The output of this activity is the mask of agricultural land of continental Africa, without making any distinction between land use types or land use pattern that can be observed. It is assumed that to develop the sampling design such a distinction is not needed in first instance. However, it is also possible and maybe advisable that distinction is made between the various land use types in the allocation of the number of sampling points. And, that the weight for the various land use types is determined based on the degree of intensity of utilization of the land for agricultural purposes.

In that case, differentiation of land use based on the rules defined in this report might provide a solution. As illustrated by the results for Tile 1, covering the area from central Africa to the Red Sea off the coast of Eritrea, the rules result in areas or zones that are clearly disjunct from each other and seem to represent distinct land use zones with their characteristic land use pattern. These zones can be easily mapped out by assigning a particular class value to each of these land use zones. Alternatively, the map of farming systems for Africa as defined by Dixon et al. (2001) could be overlaid with the masks of agricultural land produced by the SOILS4AFRICA project to provide information on the expected land use characteristics. It will not provide location specific information on the actual land use (e.g., whether a particular area is used for temporary crops, or plantation crops, for example). But, otherwise, the result is likely to be similar to the outcome of the inference rules.

Identification of areas with plantation crops, for example, is a problem with the Copernicus Global Land Cover data, like with any land cover map. We have not been able to uniquely identify plantation crops using information on the discrete land cover class and fractional cover percentage of the various cover classes. Plantation crops like sugar cane are generally classified as "Cropland" and are therefore correctly included in the mask of agricultural. But we need additional criteria to identify these (semi-)permanent plantation crops as subclass of "Cropland". A similar problem is observed with tea plantations. These are generally classified as "Forest" (mainly as "Closed forest" but also as "Open forest") and we need additional criteria related to fractional cover to identify

This project has received funding from the European Union Horizon 2020 Research and Innovation Programme under Grant Agreement 862900

these woody plantation crops. The current criteria we are using identifies the plantation crops, but it is not exclusive. There is a considerable area included in this class that does not belong to 'plantation crop' and that might be incorrectly included in the mask of agricultural land. The same will apply to vineyards, for example, and other plantation crops.

We will have to investigate the additional criteria that would allow for the unique classification of plantation crops. An option would be to use the AEZs an additional criterion. However, the best discriminating factor will be the spatial pattern and that can be only be assessed through visual interpretation of satellite imagery. In the CGLC, image texture is included in the metrics used for the pattern recognition for their land cover classification and could possibly be used for the identification of plantation crops. So far, it has not been their focus (it is a level of detail in the classification that they have not considered so far). However, the Copernicus program, might want to consider working on a further differentiation of the "Cropland" land cover class, in the same way they have done for the "Forest" class (for which they currently have three levels in the classification hierarchy). It would require redoing the training of the classification algorithms.

Another way to assure there is a proportionate representation of the various agricultural land use classes in the final selection of the sampling locations would be to assess the land use class of the tentative sampling locations by visual interpretation of the satellite imagery for those particular locations. The land use class should then be used as a criterion in the final selection of the sampling points.

We have mentioned the protected areas that should be excluded from the mask of agricultural land in Africa. This would make sense if these areas are indeed effectively protected. On the other hand, many of these protected areas are managed to some degree and there are many conservancies (wildlife reserves) that are commercially exploited and could therefore be included in the definition of agricultural land. There are also many reserves that are not effectively protected and that are actually colonized to varying degrees with even settlements having been established and villages developed in cases. Therefore, it would be advisable to scrutinize these protected areas for human settlement before excluding these areas from the mask of agricultural land in Africa and from excluding these from the sampling scheme. But this goes beyond the scope of the current activity and may be included in the preparatory activities for the sampling design. Information on the protected areas can be found on the IUCN website[4].

References

Buchhorn, M.; Smets, B.; Bertels, L.; De Roo, B.; Lesiv, M.; Tsendbazar, N.E., Linlin, L., Tarko, A. (2020): Copernicus Global Land Service: Land Cover 100m: Version 3 Globe 2015-2019: Product User Manual; Zenodo, Geneve, Switzerland, September 2020; doi: 10.5281/zenodo.3938963

Dixon J., A. Gulliver & D. Gibbon, 2001. Farming systems and poverty. Improving farmers' livelihoods in a changing world. FAO and World Bank. Rome and Washington D.C., 412 pp.

HarvestChoice; International Food Policy Research Institute (IFPRI), 2015. Agro-Ecological Zones for Africa South of the Sahara. doi: https://doi.org/10.7910/DVN/M7XIUB, Harvard Dataverse, V3.

[4] https://www.iucn.org/theme/protectedareas/about/protected-area-categories

European Commission | Horizon 2020 European Union funding for Research & Innovation

This project has received funding from the European Union Horizon 2020 Research and Innovation Programme under Grant Agreement 862900