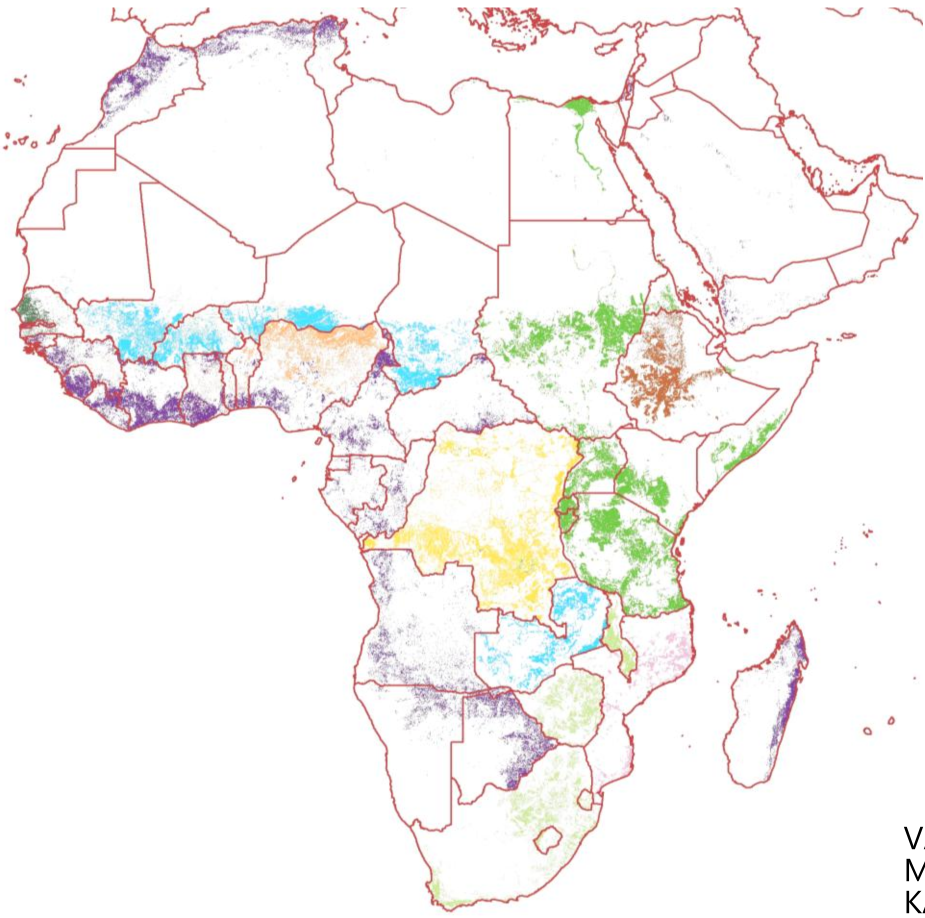
European Commission

JRC TECHNICAL REPORTS

Harmonizing and combining existing land cover and land use datasets for cropland areas monitoring at the African continental scale



VANCUTSEM Christelle
MARINHO Eduardo
KAYITAKIRE Francois
SEE Linda
FRITZ Steffen

2012

Report EUR 25570 EN

Joint Research Centre

European Commission
Joint Research Centre
Institute for Environment and Sustainability

Contact information
KAYITAKIRE Francois
Address: Joint Research Centre, Via Enrico Fermi 2749, TP 266, 21027 Ispra (VA), Italy
E-mail: Francois.Kayitakire@jrc.ec.europa.eu
Tel.: +39 0332 78 6580
Fax: +39 0332 78 5154

http://mars.jrc.ec.europa.eu/
http://www.jrc.ec.europa.eu/

This publication is a Technical Report by the Joint Research Centre of the European Commission.

JRC75684

EUR 25570 EN

ISBN 978-92-79-27160-1 (pdf)

ISSN 1831-9424 (online)

doi:10.2788/61707

Luxembourg: Publications Office of the European Union, 2012

Printed in Luxembourg

Harmonizing and combining existing land cover and land use datasets for cropland areas monitoring at the African continental scale

VANCUTSEM Christelle
MARINHO Eduardo
KAYITAKIRE Francois
SEE Linda
FRITZ Steffen

1. Introduction

Mapping cropland areas is of great interest in diverse fields, from crop monitoring to climate change and food security. Indeed, any decision making process related, directly or indirectly, with agriculture, requires precise and accurate crop extent maps in order to quantify and spatially characterize the role played by the human activity that covers the largest extent of the Earth's surface. However, in Africa, the estimation of cropland extent from remote sensing remains a challenge. Several reasons for this can be put forward: the heterogeneous nature of the agriculture, differences in crop cycles, the spatial structure of the landscape (parcel size), the spectral similarity with grassland, mainly in arid and semi-arid areas, the cloud coverage during the growing season or the inter-annual variability due to climatic events such as droughts and fallow practices.

A plethora of methods and input data have been tested and used in order to estimate the cropland extent in the continent. (i) Landsat images were used to derive cropland maps, such as the Cropland-Use Intensity (CUI) dataset (USGS) and the Africover dataset [1]. These spatially detailed maps are, in general, coherent with national statistics [2] but they are limited in their spatial coverage. Moreover, these datasets cannot be regularly updated following the methodology commonly used (i.e. visual interpretation). These issues have been addressed by (ii) the global map of cropland extent at 250m produced using multi-year MODIS data [3]. However, its global scope does not correctly account for the specificities of some regions of the globe such as Africa. (iii) Other global maps specifically dedicated to croplands were produced by the International Water Management Institute (IWMI): the global map of rainfed cropland areas (GMRCA) [4] and the global irrigated area map (GIAM) [5]. However, their coarse spatial resolution (10 km) is not suitable for national and regional applications and they present a large number of uncertainties [6]. The same resolution problem characterizes (iv) the cropland mask produced by [7] that is dedicated to agricultural lands at 10 km and combines two satellite-derived land cover maps, i.e. Boston University's MODIS-derived land cover product [8] and the GLC2000 data set [9], with an agricultural inventory. (v) More recently, existing land use/land cover datasets were combined based on expert knowledge and national statistics to produce a probability map of cropland areas [11]. However, the product is notably based on global land cover products (i.e. GLC2000, MODIS Land Cover, and GlobCover [10]) that do not focus on cropland areas and where the spatial resolution of the input remote sensing data (from 300m to 1km) is not adapted for mapping cropland areas. Moreover the map is highly dependent upon the reliability of national and sub-national statistics [11].

To improve the estimation of cropland areas at the continental scale, two complementary approaches should be considered: (i) to select the best available national and sub-national cropland maps derived, when possible, from high resolution images and combine them in order to derive a continental product and (ii) to develop procedures using high (~30m) and moderate resolution satellite data (~250m) at the national and sub-national scale to update existing cropland maps that are outdated and to create higher resolution products when only global maps exist. The objective of this study fits into the first approach and aims at combining the best existing land cover/land use datasets derived from medium resolution imagery, using a common legend based on the Land Cover Classification System (LCCS) developed by the FAO [12]. Ten datasets are harmonized and combined through an expert-based approach to derive a map of cropland areas at 250m for Africa.

The accuracy of the resulting cropland mask is compared with two recent cropland extent maps at 1km: one derived from MODIS [3] and the other derived from five existing products [11] using a validation sample of 3591 pixels of 1km² regularly distributed over Africa and interpreted using high resolution images. The validation datasets were collected through the crowdsourcing land cover validation tool called Geo-wiki, which provides a platform for the interpretation of high resolution images on Google Earth (GE) [13].

2. Data

2.1 Data sources

Ten land cover/land use products have been considered in order to produce the cropland mask (Figure 1). They are felt to be the best available crop masks for Africa to date:

- The Globcover map (2005-2006) [10], http://ionia1.esrin.esa.int/
- The SADC land cover database (CSIR, South-Africa) covering 7 countries: Lesotho, Malawi, Mozambique, South Africa, Swaziland, Tanzania and Zimbabwe.
- The Cropland Use Intensity datasets (USGS, 1988) covering 11 countries: Burkina Faso, Chad, Malawi, Mali, Mauritania, Mozambique, Niger, Somalia, Zambia, and Zimbabwe.
- The Woody Biomass map of Ethiopia (World Bank project, 2002).
- The Africover maps (FAO) covering 10 countries (2000): Burundi, Egypt, Eritrea, Kenya, RDC, Rwanda, Somalia, Tanzania, Sudan and Uganda, http://www.africover.org/
- The Democratic Republic of Congo (DRC) map [14], http://sites.uclouvain.be/enge/maps/UCL_RDC/UCL_RDC_Occupation_du_sol.html
- The land cover of Mozambique (the National Directorate of Land and Forests (DNTF) - Ministry of Agriculture, Avaliação Integrada das Florestas em Moçambique, 2007).
- The land cover map of Senegal 2005 produced by the Global Land Cover Network (GLCN, FAO) initiative, http://www.glcn.org/

- The land use / land cover (LULC) 2000 datasets produced by USGS covering 8 countries: Benin, Burkina-Faso, Ghana, Guinea, Guinee Bissau, Mali, Mauritania, Niger, Togo.
- The MODIS-JRC (Joint Research Centre, Monitoring Agricultural ResourceS (MARS) Unit, EU) crop mask derived from MODIS time series for the year 2009 over northern Nigeria and Benin [15].

It is worth noting that these datasets differ in terms of data source, resolution, methodology, geographical extent, and time interval used for deriving the land cover/land use dataset (Table 1).

Table 1. Description of each land cover/land use dataset in terms of reference, data source and resolution, and time interval

	Globcover	LC Mozambique	SADC	CUI	LULC 2000	Woody Biomass	DRC map	Africover	GLCN Senegal	MODIS-JRC
Reference	Defourny et al. 2009	DNTF	CSIR	USGS	USGS	World Bank	Vancutsem et al. 2009	FAO	FAO	Vancutsem et al. 2011
Data source	MERIS	Landsat	Landsat 7 TM	Landsat 5 MSS	Landsat	Landsat	SPOT VEGETATION	Landsat 7 TM	Landsat	MODIS
Data resolution	300m	100m	30m	30m	2km	30m	1km	30m	30m	250m
Time interval	2005-2006	2007	1990-1995	1986-1988	2000	2000	2000	1995 - 2002	2003-2005	2009

2.2 Validation datasets and the Geo-wiki tool

The validation datasets used in this study are based on the interpretation of high resolution images on Google Earth (GE) through the crowdsourcing land cover validation tool called Geo-wiki [13]. Geo-wiki has built up a global network of volunteers that are helping to improve the quality of global land cover maps. Volunteers are asked to determine the land cover type from a simple legend based on Google Earth imagery and their local knowledge. Their input is recorded in a database, along with uploaded photos, which are currently being used to create an improved global land cover map. Several Geo-wiki variants are available, each focusing on different land cover types (e.g. biomass.geo-wiki.org, urban.geo-wiki.org [16]) including cropland (http://agriculture.geo-wiki.org). In the agricultural version of Geo-Wiki, users are asked to provide the percentage of cropland that they can see from Google Earth imagery at intervals of 10% (0, 1-10,…,90-100), as well as the confidence associated with their assessment. In addition, users can specify if a high resolution image was available and its acquisition date.

Two datasets were used: (i) IIASA experts validated pixels of 1km² resolution located at each latitude/longitude intersection point across Africa (2942 pixels), and (ii) pixels of 250m² resolution were interpreted by MARS (JRC, European Commission) experts (649 pixels) every 0.36 degree, covering mainly Niger and northern Nigeria.

3. Methodology

The methodology follows three steps: (i) harmonization of the ten existing land cover/land use datasets in order to have comparable products, (ii) selection of the cropland classes to be used for the crop mask, (iii) identification of the best product for each country in order to spatially combine them and derive a map of cropland areas at 250m for Africa. This section describes these steps in more detail as well as the validation of the resulting cropland mask.

3.1 Harmonization of the datasets

The land cover/land use products used in this study have different sources and consequently different projections, formats, resolutions, and legends. Therefore the following processes were applied prior to combining the different datasets:

1- Legend harmonization: Given the heterogeneity of the products, this is the most critical step in the process. The Land Cover Classification System (LCCS) developed by the FAO aims to analyze and cross-reference regional differences in national land cover descriptions [12]. Some of the products have used the LCCS in their development (Globcover, DRC map, Africover, GLCN and the MODIS-JRC dataset) while others have not. Therefore, those products that have not adopted the LCCS differ in how they define agriculture (i.e. land cover description, land use intensity) since the aims of these products differ. In this study, a common legend with five cropland classes in accordance with Globcover standards has been adopted. These are described in Table 2 using the LCCS. All the cropland classes in each of the products were then mapped to this legend (Table 3).

For those products that used the LCCS (Globcover, DRC map, Africover, GLCN and the MODIS-JRC dataset) in their development, the conversion of the legend was straightforward. For the CUI dataset, the conversion was more complex since the product describes five levels of agricultural land use intensity (0-5%, 5-30%, 30-50%, 50-70%, 70-100%) using Landsat images from 1988 that do not correspond to the current reality because of the intensification of agriculture that has occurred since this period. A visual analysis of the product in comparison with recent high-resolution imagery available on Google Earth was then required. Therefore, based on an expert-based visual analysis, the CUI classes have been converted as follows (Table 3):

- Levels 1 and 2 of the CUI (50-100%) become “Cultivated and managed areas (70-100%)”
- Level 3 (30-50%) becomes “Mosaic cropland (50-70%)/vegetation”
- Level 4 (5-30%) becomes “Mosaic vegetation/cropland (20-50%)”
- Level 5 (0-5%) is considered as natural vegetation

Table 2. Legend description for cropland classes with the LCCS codes and classifiers

User Legend (Globcover)	LCCCode	LCCClassifier
Cultivated and Managed areas	0003	A11
Post-flooding or irrigated croplands	11491//11495//11500//11499	A1XXXXXXD3//A2XXXXXXD3//A3XXXXXXD3//A3XXXXXXD2
Rainfed croplands	11494 // 11490 // 11498	A2XXXXXXD1 // A1XXXXXXD1 // A3XXXXXXD1
Mosaic cropland (50-70%) / vegetation (grassland/shrubland/forest) (20-50%)	0003 / 0004	A11 / A12
Mosaic vegetation (grassland/shrubland/forest) (50-70%) / cropland (20-50%)	0004 / 0003	A12 / A11

Table 3. Translation of the ten legends into the common legend

MapCode	User Legend	Globcover	LC Mozambique	SADC	CUI	LULC 2000	Woody Biomass	DRC map	Africover AG	GLCN Senegal	MODIS-JRC
10	**Cultivated and Managed areas**	10	1FC (field crop)	Cultivation		Agriculture					Cultivated and Managed areas
11	**Post-flooding or irrigated croplands**	11, 12, 13	3AC (cultivated aquatic or regularly flooded area)		A1, A2, A3, DI1, DI2, DI3, RC1, RC2, RC3	Irrigated agriculture	Cultivated Land; Irrigated		15, AG-16, AG-17, AG-10,AG-11,AG-12,AG-18,AG-19,AG-20,AG-23,AG-24	1H-p-de, 1arV-g-Ir 1H-g-Ir-suc, 3HH-g, 1H-p-Ir, 1H-g-Ir, 3H-p	
14	**Rainfed croplands**	14, 15, 16	1TC (Tree crops), 1SC (Shrub crops)	Plantation	HB1, HB2, HB3, S1, S2, U1, U2	Plantation	Cultivated Land; Rainfed & shifting cultivation, Plantation forest	13 (Agriculture)	AG-1,AG-2,AG-3, AG-21,AG-22,AG-4,AG-5,AG-6,AG-7,AG-8,AG-9	1H-g, 1H-p-ls, 1H-p+A, 1H-p+A-ls, 1H-g+A, 1H-p, 1APf-g-2, 1AV-p, 1APf-p-1, 1AV-g, 1APf-g-1	
20	**Mosaic cropland / vegetation**	20, 21, 22	1CXF (Shifting cultivation with forest + other vgt)		S3, U3, HB3					1APf-g-1, 1APf-g-2, 1AV-p, 1H-g-Ir, 1H-g, 1H-g+A, 1H-p-de, 1H-p-Ir/1AV-p, 1H-p-Ir, 1H-p, 1H-p=A	
30	**Mosaic vegetation / cropland**	30, 31, 32	2FXC (Forest with shifting cultivation)		S4, U4, HB4, A4, RC4, DI4			7 (rural complex), 9 (mozaic steppic sav-crops)		1H-p-Ir, 2arTO, 3H-p, 1H-p+A	

2- Conversion from feature to raster: Vector datasets such as Africover, CUI and SADC are converted into a 250m resolution raster using the “maximum area” criteria, i.e. the feature with the largest area in the cell yields the attribute assigned to that cell.

3- Reprojection: Datasets using other projections (the CUI dataset, the LULC2000 product and the land cover of Mozambique) were reprojected to the Geographic projection (WGS84).

4- Geometric correction: A spatial shift was observed between the CUI and the other products and corrected accordingly.

5- Resampling: All datasets have been resampled at a 250m resolution. The reason is that this resolution is compatible with MODIS time series, which are the highest resolution images used for the monitoring of agriculture in Africa.

3.2 Selection of cropland classes

By default, all cropland classes with a majority of cropland areas (>50%) were integrated into the crop mask. For all classes with a minority of crops (20-50%), a visual analysis was undertaken by several experts who compared the dataset with high resolution images on GE. Based on this analysis, this class was integrated in the final cropland mask for the Globcover dataset only. Indeed, for some countries in equatorial areas where only mosaic classes are available and/or where cropland areas are mixed with forest, it was preferred to take this class into account in order to avoid underestimating cropland areas in these regions.

For countries where the information on irrigated crops was available, irrigated and rainfed crops have been contrasted. Therefore, three crop masks were produced: (1) one with irrigated and rainfed cropland areas, (2) one containing only irrigated crops, and (3) one with only rainfed crops.

3.3 Spatial combination of the datasets

As overlaps between datasets exist, a priority ranking was determined in order to combine them. Datasets were compared using high resolution images (GE) and the best product has been selected on a case by case basis. In cases where the choice was not obvious for the expert(s) because of product similarity, the following rules were applied:

- Priority was given to the product with the highest spatial resolution;
- If two products of the same resolution exist, priority was given to the most recent product.

In some cases, when two datasets were complementary, the cropland classes of both products were spatially combined.

3.4 Validation

The validation is based on two validation samples, one with 2942 pixels of 1km² covering Africa and another partially covering Niger and Nigeria with 649 pixels of 250m². These samples were visually interpreted by experts using high resolution images through the crowdsourcing land cover validation tool called Geo-wiki [13]. The three products were validated against this reference: (i) the crop mask obtained by the above described methodology and hereafter referred to as MARS-JRC and two recent cropland products at 1 km spatial resolution, (ii) the global cropland extent map derived from MODIS [3] and (iii) the cropland extent product of Africa derived from five existing products [11]. The last two last datasets provide the percentage of cropland while the first is binary in nature. The combined validation of these three products aims to evaluate the improvement, if any, that the MARS-JRC crop mask generates.

In addition, a limited sample of 179 pixels covering Niger and Nigeria has been used in order to assess the extent of the discrepancies between experts. In each of these randomly selected pixels, the vegetation coverage has been estimated by two overlapping experts. This step of the analysis is considered to be important if one aims to understand the scope of the validation in the crowdsourcing environment.

4. Results

4.1 Crop mask and sources

The resulting crop mask is delivered in Geographic projection (WGS84) at a spatial resolution of 0.00208333 degrees. Figure 1 shows the source that has been used for each pixel. Of the 44 countries with significant cropland areas, 23 are covered by a dataset derived from Landsat images, 3 are covered by regional land cover products and 18 by a global land cover product (Globcover).

For 8 countries (Burkina Faso, Chad, Mali, Mauritania, Mozambique, Niger, Tanzania and Zimbabwe), two or three Landsat-based products were available and it was thus necessary to compare carefully the different cropland maps. Based on the knowledge of experts and high resolution images (GE), either the best product was selected or the classes derived from different products were spatially combined.

For Mozambique, three products were available (Figure 2): the Cropland Use Intensity dataset (USGS 1986-1988), the SADC land cover database (CSIR, South-Africa) and the land cover of Mozambique (the National Directorate of Land and Forests (DNTF). According to the experts and the visual analysis with GE, the SADC product was missing cropland areas mainly in the Nampula region (North-East) and in the Manica region (from Manica to Chimoio, Center of the country) which disqualified the product for selection. However, the choice between the two other products was not obvious. An overestimation of the cropland areas has been observed in the CUI dataset whereas some crops were missing in the DNTF dataset, in particular in the regions of Tete and Nampula. As no product seemed better than the other one (and the best probably is somewhere between the two products), priority was given to the most recent product, i.e. the DNTF dataset.

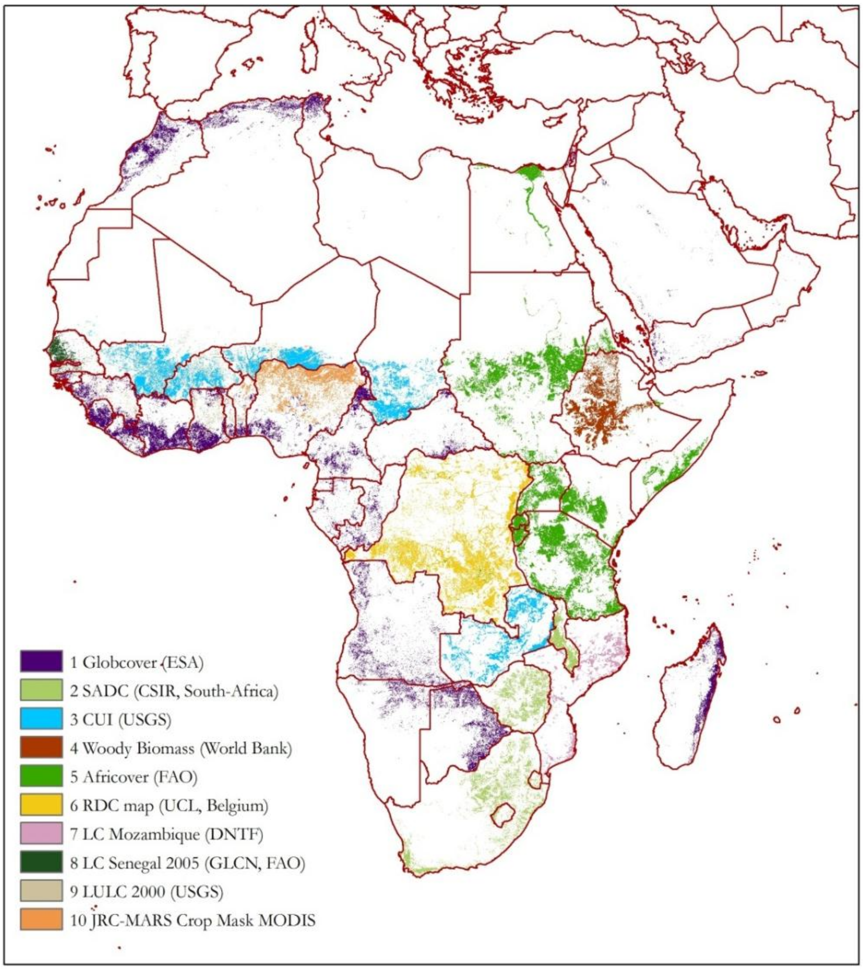


Figure 1. Selected sources for each pixel of the MARS-JRC crop mask

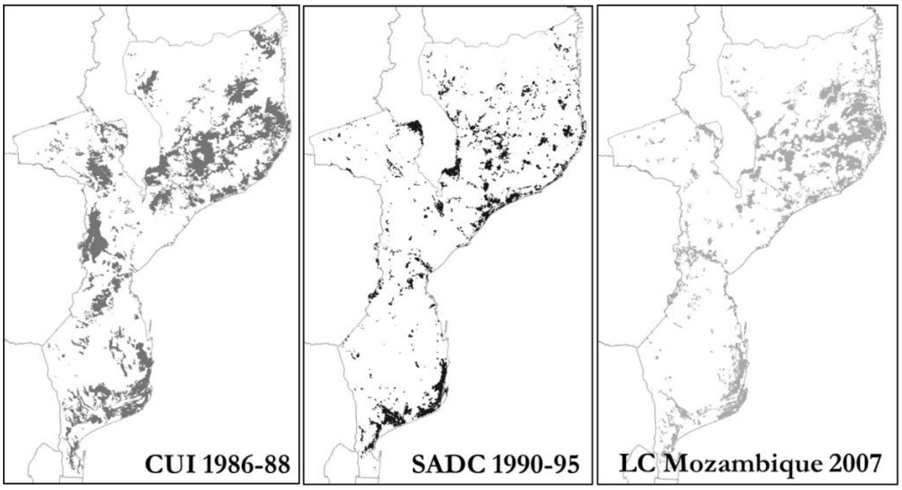


Figure 2. Cropland maps of Mozambique: the Cropland Use Intensity dataset (USGS 1986-1988), the SADC land cover database (CSIR, South-Africa) and the land cover of the National Directorate of Land and Forests (DNTF).

For West African countries covered by the CUI dataset and the LULC 2000 (Burkina Faso, Chad, Mali, Mauritania, and Niger), the cropland classes of both products were found to miss important cropland areas. Fortunately, a meticulous analysis of both products showed that the nature of the omissions is different and that they appear to be complementary. In the CUI dataset (1986-1988) the images used are outdated especially for agriculture as cultivated areas expanded considerably in these areas during the last 20 years. In the LULC dataset, omission errors have also been observed most likely due to the coarse resolution (2 km) of the available product (the original product is at 30m but is not available). Therefore, both datasets have been combined and a better product has been obtained. Figure 1 shows an example of this spatial combination for South-West Niger.

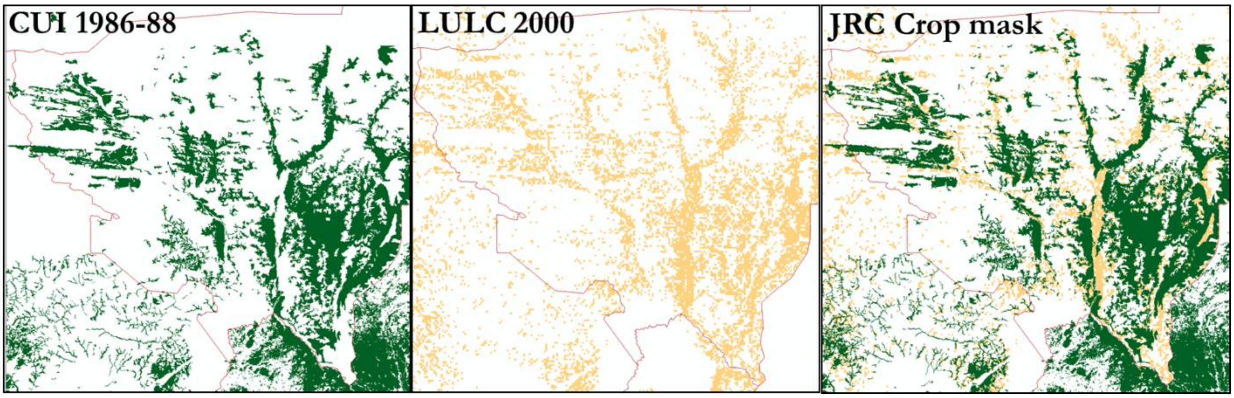


Figure 2. Combination of Cropland maps in western Niger: the Cropland Use Intensity dataset (USGS 1986-88), the LULC (USGS 2000) product, and the resulting MARS-JRC.

In Tanzania, the Africover dataset was preferred to the SADC dataset, mainly because important cropland areas were missing in the SADC product, notably in the region of Dodoma between Lake Rukwa and the border with Zambia (South-West), and in the North-East of Lake Tanganyika close to the border with Burundi.

For Zimbabwe and Malawi, based on comparison with high resolution images (GE), the SADC map was preferred to the CUI dataset as omission and commission errors were observed in the latter. For instance in Malawi, few cropland areas were identified in the Lilongwe districts (around the capital) whereas GE imagery clearly shows cultivated areas. On the other end of the spectrum, the Dzalanyama Forest (in the West of Lolongwe) and the Majete Wildlife Reserve (in the South) were partially identified as crops.

In Democratic Republic of Congo, the Africover dataset seemingly misses vast cropland areas. Some of those areas are labeled as pure and mixed cropland classes in the DRC map based on SPOT VEGETATION time series [14] and are combined within the cropland areas of the Africover dataset.

For the countries where Globcover and only one Landsat-based product were available, the Landsat-derived product was selected. These countries are (i) Burundi, Egypt, Eritrea, Kenya, Rwanda, Somalia, Sudan, and Uganda, covered by Africover, (ii) South-Africa, covered by the SADC dataset, (iii) Senegal, covered by the GLCN dataset, (iv) Ethiopia with the Woody Biomass product and (v) Zambia, covered by the CUI dataset.

In Nigeria, the Globcover and another product derived from MODIS data were available (MODIS-JRC, [15]). The second product was preferred to the first because (i) of the better spatial resolution and (ii) of the classification methodology, which focuses on agricultural lands.

Finally, the crop mask derived from Globcover (classes 10 and 20, plus class 30 for the equatorial countries) was used for the remaining countries.

4.2 Validation

4.2.1 Qualitative assessment

The first step in the validation process of the three cropland products, i.e. MARS-JRC and the maps produced by Pittman et al. [3] (referred to hereafter as Global Croplands) and the IIASA African cropland product beta version of Fritz et al. [11] (referred to hereafter as the IIASA product), is an expert visual assessment assisted by high resolution images available on GE. For this purpose, vast territories, randomly distributed throughout Africa, have been visually analysed, from which three regions, i.e. Niger/Nigeria, Sudan, and South Africa (Figure 4) have been selected to be the focus of this section. These regions are characterized by (i) substantial discrepancies between the three cropland maps, (ii) cultivated areas that can be spotted with high confidence levels using GE imagery and (iii) a large spectrum of agricultural intensity, from extensive farming in Sudan to intensive farming in South Africa. The observed differences between the three products are:

- For Nigeria (1) and Sudan (2), the IIASA product appears as the inverse of the two other products from what can be observed on GE images. For instance, in Nigeria, the west Tangaza Forest Reserve, characterized by a savannah land cover (see A at Figure 4), has been classified as crops in the IIASA product while the surrounding cropland areas have been classified as natural vegetation. The use of Globcover and GLC2000, which do not have agriculture as their specific focus, in the IIASA product, may explain these mis-interpretations. The same is not observed in the other two products that correctly exclude/include the savannah and surrounding croplands in the crop mask. In Sudan, the IIASA product and the MARS-JRC product are based on the Africover dataset. Therefore we would have expected a similar result to the MARS-JRC crop mask. However, the product is missing cropland areas in the North and overestimates crop areas in the South.

-The Global Croplands and the MARS-JRC products have similar patterns especially for regions (1), and (2) but the Global Croplands product shows a lower percentage of crops.

-In Sudan, the MARS-JRC crop mask correctly depicts cropland areas south of the river, in particular the linear structures in the South of Umm Ruwaba, i.e. the succession of crop and fallow areas (see B at the Figure 4). These areas have not been detected by the two other products.

-In South-Africa, the MARS-JRC and Global Croplands crop masks identify the Hoopstad region (in light orange on GE, C at the Figure 4) which is the richest maize-production district in South Africa, while it is missed by the IIASA product. However, the savannah area in the south of Hoopstad (in green on GE, D at the Figure) and the cropland areas in the South of

-dark orange, E at the Figure 4) are not well identified on the Global Croplands product as opposed to the MARS-JRC map and partially in the IIASA product.

Large discrepancies and sometimes inverted classifications have been observed between the three products. The largest errors primarily result from the use of global land cover products, where the thresholds are optimized for multi-class identification problems and are not specifically dedicated to agriculture. Moreover, coarse resolution data, which is suitable for global and continental products, performs worse than Landsat-derived datasets. Therefore, the use of a unique satellite image processing chain at the global scale, even when dedicated to agriculture like that of Global Croplands, cannot depict all the cropland areas of the globe with the same intensity, even if the spatial patterns are correct. Finally, the calibration using national statistics may induce errors whereas the satellite-derived map is correct (e.g., Africover in Sudan).

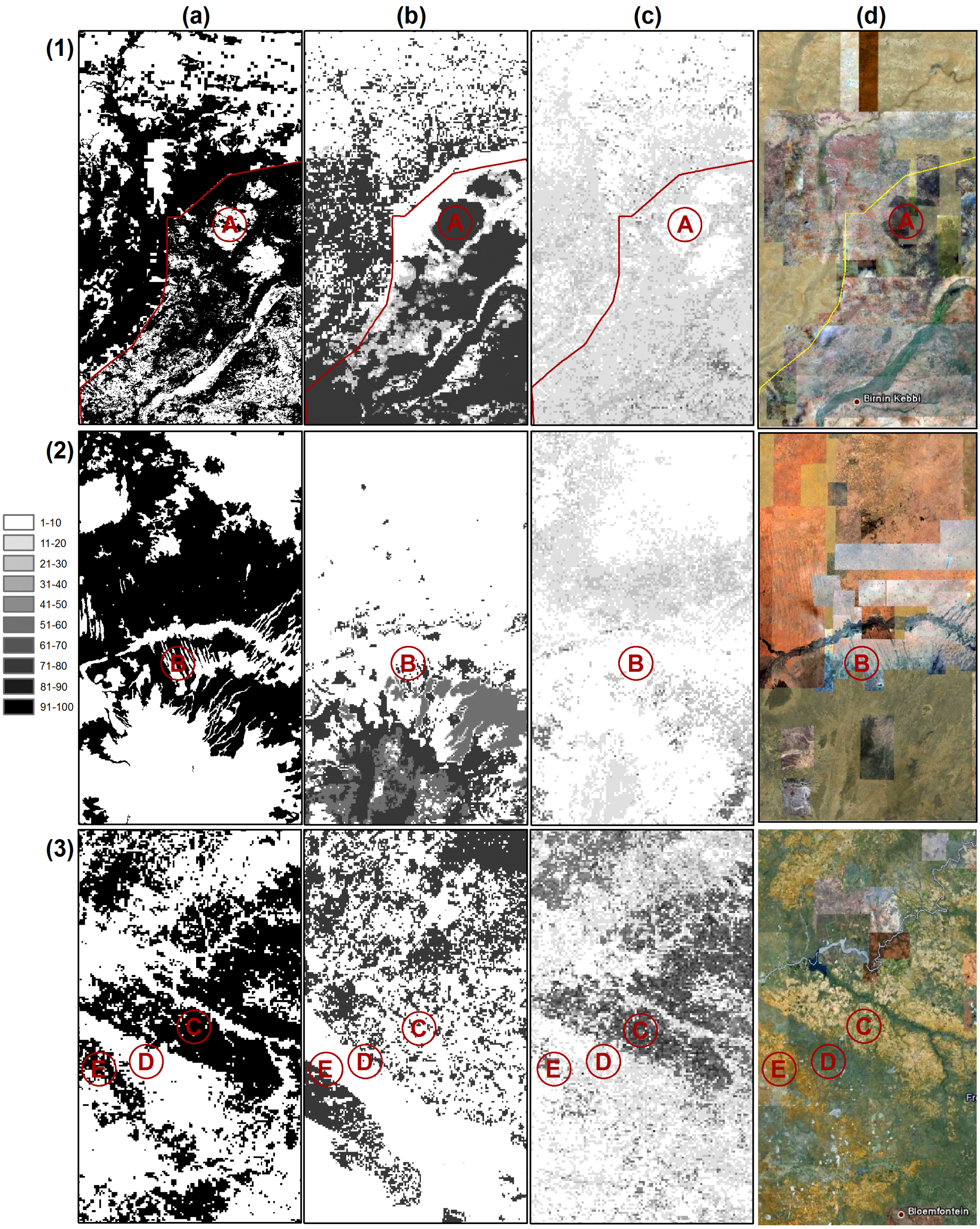


Figure 4. Comparison of the three crop masks: (a) MARS-JRC, (b) IIASA product and (c) Global Croplands with (d) GE images for three regions: Niger/Nigeria (1), Sudan (2), and South Africa (3). The upper left coordinates for these three regions are respectively (lat, long): (3.6,14.5), (30.7,13.9), and (25.4,-26.9).

4.2.2 Consistency between experts

Before proceeding to the quantitative accuracy assessment, which takes as reference the expert interpretations coming from the Geo-wiki tool, one has to evaluate the extent of the agreement between experts. Indeed, high accuracy levels are expected if and only if experts agree between them. Otherwise, divergences between the views of the experts should be taken into account in the interpretation of the results of the accuracy assessment.

For the Niger-Nigeria validation window, the sampling was defined in order to have an overlap of 180 pixels between two experts, where both experts were confident regarding their interpretation of 130 pixels. For the remaining 50 pixels, at least one of the experts was unsure or did not give any interpretation. The percentage of agreement between both experts was computed in two ways: (i) for each category of percentage of croplands (% of agreement for 1 class), and (ii) given the difficulties of determining percentages in a 250m box, by accepting confusions between neighbouring percentage classes (% of agreement for 3 classes).

Figure 5 shows that higher agreement levels between experts are observed for classes with no or almost full agricultural coverage (90-100%), with respectively 83.1% and 45.2% of agreement. It is worth noting that those classes are the ones most commonly represented - together they include 78.5% of the assessed pixels - as a consequence of the usual clustering of agricultural areas. However, agreement strongly deteriorates for classes with partial agricultural land coverage. Although the agreement estimates for those classes are quite unreliable given the fact each one has only 5 to 10 overlapping pixels, this can be taken as evidence of how difficult it is for a human being to assess the percentage of coverage.

This is confirmed by an increase in the percentage of agreement when confusions between neighboring classes are not taken as errors (dark bars). Although higher agreement levels are still observed for classes with high or low agricultural coverage, a significant increase in the agreement is observed for all intermediate classes.

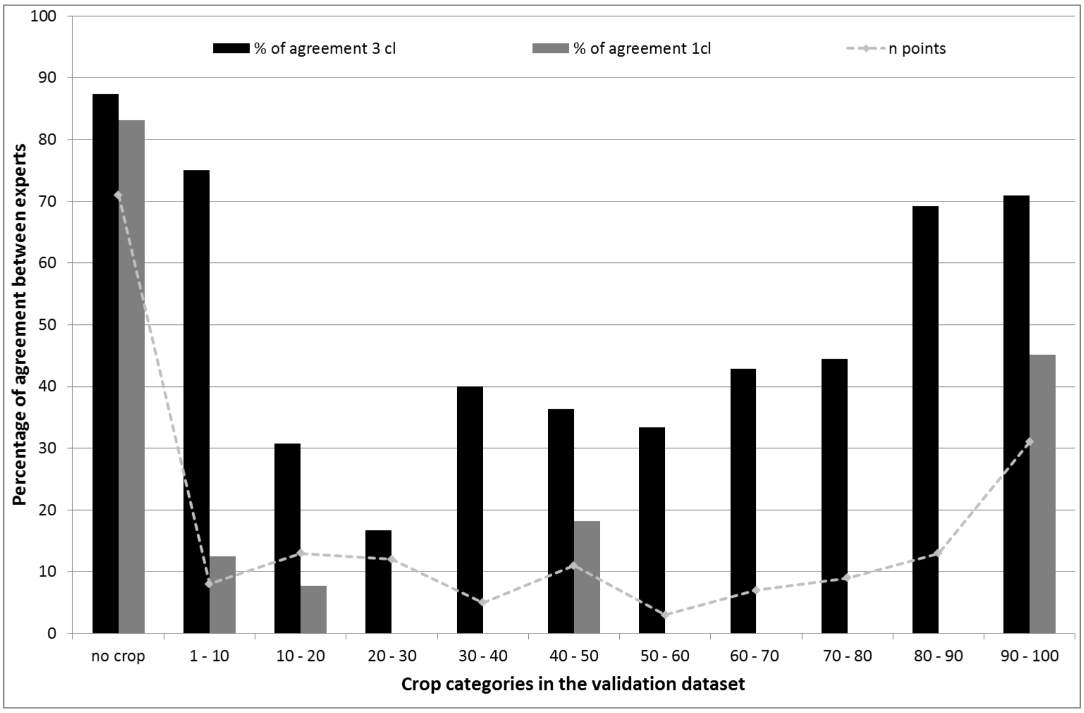


Figure 5. The percentage of agreement between two experts for each category of crops taking into account the category concerned only (% of agreement for 1 class in grey) but also the neighbouring classes (% of agreement for 3 classes in black), and the number of pixels interpreted by both experts for each category (dashed grey line).

Table 4 presents the agreements and disagreements between experts and for each category. For 13 pixels, the disagreement is high as a difference of at least 50% is observed between the percentages given by the two experts. These results highlight the difficulty to map different agricultural land use intensities, even with labour intensive procedures such as visual interpretation, in particular in semi-arid areas. Despite the use of high resolution imagery and visual interpretation, experts only agree on the extreme intensity values. When considering only two classes of presence (more than 50%) and absence (less than 30%), the level of consistency between experts is respectively 65.1% and 70.2%, which is acceptable for the validation process if we consider only these classes of low and high crop densities (see confusion matrixes in Table 5).

4.2.3 Accuracy assessment

The percentage of pixels detected as crop for the three crop masks is plotted in Figure 6 against each category of crop coverage intensity for the African validation datasets (2942 pixels). The more accurate a crop mask is, the closer it will be to diagonal. Indeed, within the area of each crop coverage intensity class, the ideal crop mask is expected to provide the same percentage of crop coverage. However, given the inconsistencies between experts for the middle range classes and the limited number of validation pixels, especially for the Niger-Nigeria area, notable deviations are anticipated in this part of the graph.

Table 4. Number of agreements between experts for each category of crops and the percentage of agreement between both experts for each category of crops taking into account the category concerned only (% of agreement for 1 class) but also the neighbouring classes (% of agreement for 3 classes).

	no crop	1 - 10	10 - 20	20 - 30	30 - 40	40 - 50	50 - 60	60 - 70	70 - 80	80 - 90	90 - 100
no crop	59										
1 - 10	3	**1**									
10 - 20	2	2	**1**								
20 - 30	3	1	1								
30 - 40				1							
40 - 50	2		2	1	1	**2**					
50 - 60			1			1					
60 - 70				2	1						
70 - 80		1	2			1		3			
80 - 90	1		1				1	1	1		
90 - 100	1		1	3	2	1			1	8	**14**
difficult to decide	27	2	2	3	1	2	3	3		1	5
sum	71	8	13	12	5	11	3	7	9	13	31
% agreement 1 cl	83,1	12,5	7,7	0,0	0,0	18,2	0,0	0,0	0,0	0,0	45,2
% agreement 3 cl	87,3	75,0	30,8	16,7	40,0	36,4	33,3	42,9	44,4	69,2	71,0

Over Africa, the MARS-JRC crop mask performs remarkably better than the 2 other products. Indeed, while the IIASA Africa and the Global Croplands products appear to underestimate crop areas in all pixels having more than 30% of agricultural coverage (i.e. the point from which the lines start to steadily diverge from the diagonal), the MARS-JRC product better characterizes the increase on cropland surfaces. The deviation from the diagonal indicates a considerable underestimation, greater than 70% in the 90-100% crop coverage class, for both the IIASA Africa and the Global Croplands crop masks, while the MARS-JRC product underestimates this crop cover class by roughly 30%. However for the category “no cropland”, the 3 masks have a similar percentage of crops, i.e. between 3.5% and 7.7%.

The analysis of the consistency between experts presented in the last section motivates the next step in our assessment where confusion matrixes were calculated for the 3 assessed crop masks in the cases where the agricultural land intensity is converted to a binary variable based on the following 3 thresholds: 1%, 30% and 50% (Table 5).

The MARS-JRC product appears to markedly better identify crops than the other products regardless of the threshold adopted. For example, the MARS-JRC product is able to detect between 46.7% to 65.1% of the cropland areas whereas the IIASA product detects between 32.8% to 39.4%. The Global Croplands product detects only 10.1% to 16.1% respectively with the thresholds 30 and 50% but it detects 99.3% of the cropland areas when aggregating all the crops from 1 to 100%. This happens because the majority of the values for the Global Croplands product vary between 1 and 10%. On the other hand, the IIASA and Global Croplands products suffer from smaller commission errors except for the Global Croplands products and the first category.

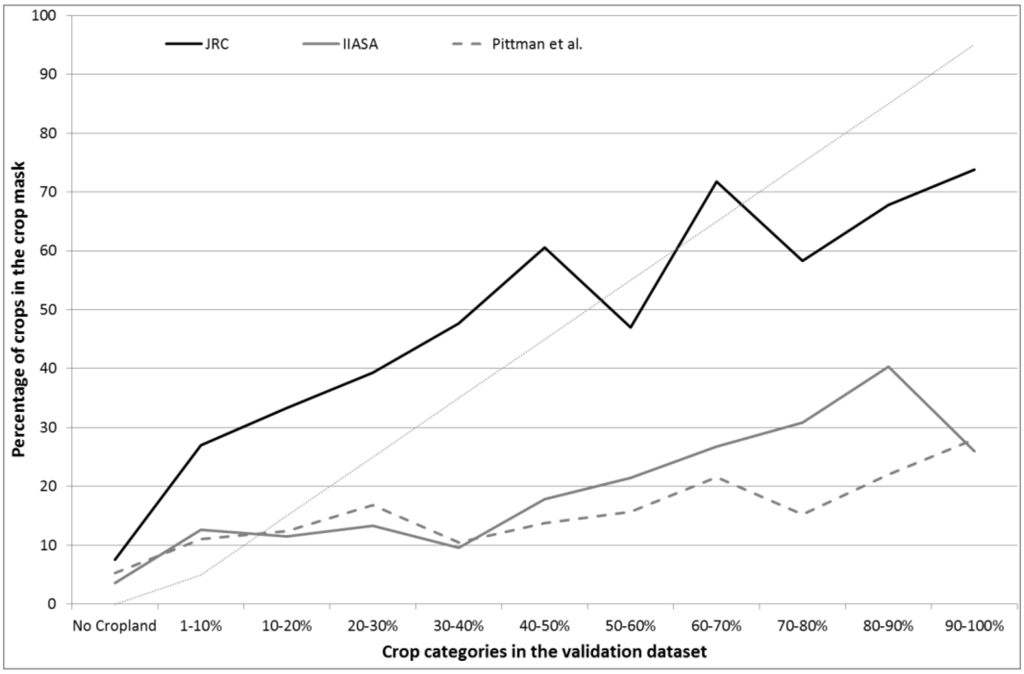


Figure 6. Agreement between the three cropland extent products (dark line for JRC, grey line for IIASA, dotted grey line for Global Croplands) and the validation sample for each category of crops based on the African dataset (2942 pixels).

Although less validation pixels are available over the area covering Niger and Nigeria, some interesting results can be observed from the confusion matrices for the 3 types of aggregation (1-100%, 30-100%, 50-100%) (Table 6). A threshold of no crops - 30 percent was chosen but it is not easy to detect cropland below this threshold using medium resolution remote sensing. However, particularly in areas of low population density and in areas of shifting cultivation, these lower percentages will occur. Future products, which are based on classification of Landsat [17], should be able to detect these smaller scale cultivation patterns and it will be possible to lower this threshold for the validation.

The IIASA crop mask performs better than for Africa as a whole with better percentages than the MARS-JRC product for the “crops 1-100%” class, and presents similar results for the second category (30-100%). However the MARS-JRC product better identifies crops than the other products for the third category (50-100%) and always performs better for the “no crop” class. The Global Croplands product tends to underestimate the percentage of crops for the categories above 30 and 50% (9.8% and 3%) but less contamination errors are observed.

Table 5. Confusion matrices based on the African dataset (2942 pixels) for the three cropland extent products, i.e. MARS-JRC, Fritz et al., and Global Croplands products, for three types of aggregation: (i) cropland present (crops 1-100%) versus cropland not present, (ii) cropland with a 30% threshold vs no cropland below 30%, and (iii) cropland with a 50% threshold vs no cropland below 50%.

Ref \ JRC	No crop	Crops 100%
No crop	**92,41**	7,59
Crops 1-100%	53,26	**46,74**

Ref \ IIASA	No crop	Crops 1-100%
No crop	**93,51**	6,49
Crops 1-100%	67,21	**32,79**

Ref \ Pittman	No crop	Crops 1-100%
No crop	**0,68**	99,32
Crops 1-100%	0,72	**99,28**

Ref \ JRC	No crop	Crops 100%
No crop - 30%	**89,96**	10,04
Crops 30-100%	38,21	**61,79**

Ref \ IIASA	No crop - 30%	Crops 30-100%
No crop - 30%	**94,10**	5,90
Crops 30-100%	65,71	**34,29**

Ref \ Pittman	No crop - 30%	Crops 30-100%
No crop - 30%	**97,72**	2,28
Crops 30-100%	83,93	**16,07**

Ref \ JRC	No crop	Crops 100%
No crop - 50%	**88,64**	11,36
Crops 50-100%	34,85	**65,15**

Ref \ IIASA	No crop - 50%	Crops 50-100%
No crop - 50%	**94,36**	5,64
Crops 50-100%	60,61	**39,39**

Ref \ Pittman	No crop - 50%	Crops 50-100%
No crop - 50%	**99,26**	0,74
Crops 50-100%	89,90	**10,10**

Table 6. Confusion matrices based on the Niger-Nigeria dataset (649 pixels) for the three cropland extent products, i.e. MARS-JRC, Fritz et al., and Global Croplands, for three types of aggregation: (i) cropland present (crops 1-100%) versus cropland not present, (ii) cropland with a 30% threshold vs no cropland below 30%, and (iii) cropland with a 50% threshold vs no cropland below 50%.

JRC	No crop	Crops 100%
No crop	**80,09**	19,91
Crops 1-100%	41,49	**58,51**

IIASA	No crop	Crops 1-100%
No crop	**72,85**	27,15
Crops 1-100%	28,22	**71,78**

Pittman	No crop	Crops 1-100%
No crop	**1,36**	98,64
Crops 1-100%	0,00	**100,00**

JRC	No crop	Crops 100%
No crop - 30%	**74,05**	25,95
Crops 30-100%	36,42	**63,58**

IIASA	No crop - 30%	Crops 30-100%
No crop - 30%	**67,82**	32,18
Crops 30-100%	35,84	**64,16**

Pittman	No crop - 30%	Crops 30-100%
No crop - 30%	**95,16**	4,84
Crops 30-100%	90,17	**9,83**

JRC	No crop	Crops 100%
No crop - 50%	**72,17**	27,83
Crops 50-100%	30,37	**69,63**

IIASA	No crop - 50%	Crops 50-100%
No crop - 50%	**59,02**	40,98
Crops 50-100%	39,26	**60,74**

Pittman	No crop - 50%	Crops 50-100%
No crop - 50%	**97,25**	2,75
Crops 50-100%	97,04	**2,96**

5. Conclusion

Recognizing the value of a reliable and harmonized crop mask that covers the entire African continent, the objectives of this study were to (i) consolidate the best existing land cover/land use datasets, (ii) adopt the Land Cover Classification System (LCCS) and (iii) assess the final product. Ten datasets were harmonized and combined through an expert-based approach and the derived map of cropland areas at a resolution of 250m covering the whole of Africa has been presented.

For the majority of the countries affected by rainfall variability and that are of interest for food security monitoring, the final cropland map includes the best Landsat-derived products and in some cases a spatial combination of several ones, which is a great advantage compared to existing continental products. For some regions the maps used are quite outdated as there is no recent dataset available but the objective is to update it regularly when datasets become available.

The crop mask obtained (MARS-JRC) was compared with two recent cropland products at 1 km spatial resolution: a global cropland extent map derived from MODIS [3] and a cropland extent product of Africa derived from five existing products [11]. The three products were validated against two validation samples, one with of 2942 pixels covering Africa and another partially covering Niger and Nigeria with 649 pixels, which were based on the expert visual interpretation of high resolution images using the agriculture.geo-wiki.org tool.

In addition, some pixels were validated by two experts in order to assess the consistency between them. The analysis shows acceptable levels of consistency between experts in pixels having large and low crop density but also highlights consistency problems for the classes with intermediate cropland concentrations. However, it is worth noting that the analysis is based on only 130 pixels and should be undertaken again using a larger dataset in the future.

The comparison of the resulting crop mask with existing products shows that the MARS-JRC crop mask has a greater agreement with the expert validation dataset, in particular for the Africa dataset with cropland above 30%. In particular, the other two datasets underestimate cropland intensity in areas of high density. The reason for the underestimation by these 2 other products compared to the MARS-JRC crop mask might be related to the fact that both products have been calibrated with FAO statistics at the national level and FAO might be underestimating agricultural areas. Alternatively, the way that these products have been calibrated could be altered to increase the cropland extent. For example, cropland definitions vary from 60% to 100% so in the case of the IIASA product, an average of 80% was chosen for crop distribution. If a lower average had been assumed, a greater amount of agricultural land would be distributed until the area matches the FAO statistics. This would result in a larger overall cropland extent.

This study highlights the importance: (i) of using regional and national land cover/land use datasets instead of global datasets to identify cropland areas in Africa, (ii) of using expert knowledge and high resolution images to select the best datasets available, and (iii) of being cautious when using statistical data that can sometimes be inaccurate. Moreover, it also shows that cropland extent maps derived from medium resolution time series should be encouraged as it offers the possibility of generating consistent large area crop cover maps with a higher update frequency than higher spatial resolution data. However, it requires a regional approach adapted to the specificities of each region in order to provide percentages of crops that are more consistent with reality.

As the combined cropland masks are based on various input datasets, the resulting product may present spatial inconsistencies. Indeed, the “real” spatial resolution of the datasets used, the thematic content, and the data and methodologies used in their creation, are variable from one dataset to another. However the use of higher resolution datasets, and the common resolution and legends adopted have reduced these inconsistencies.

Acknowledgments

The authors gratefully acknowledge people or institutions who contributed to the improvement of the crop mask by providing data or their knowledge of the field: Oscar Rojas (FAO), Olivier Le fay (EU delegation in Niger), the Ministry of Agriculture in Mozambique, Giulia Conchedda (FAO), Felix Rembold (EU-MARS), Philippe Mayaux (EU-GEM) and Gray Tappan (USGS).

The authors also acknowledge all the experts that contributed to the interpretation of the validation pixels on the agriculture.geo-wiki: Markus Tum, Ian McCallum, Luke Burns, Adriana Gomez, Rene Gommes, Giancarlo Pini, Ferdinando Urbano, Michele Meroni.

References

1. FAO, 1998. Land cover and land use, The FAO AFRICOVER Programme, URL: http://www.africover.org/.
2. Fritz, S.; See, L.; Rembold, F. Comparison of global and regional land cover maps with statistical information for the agricultural domain in Africa. *International Journal of Remote Sensing*, **2010**, 31, 2237-2256.
3. Pittman, K.; Hansen, M.C.; Becker-Reshef, I.; Potapov, P.V.; Justice, C.O. Estimating Global Cropland Extent with Multi-year MODIS Data. *Remote Sensing*, **2010**, 2, 1844-1863.
4. Biradar, C.M.; Thenkabail, P.S.; Noojipady, P.; Li, Y.; Dheeravath, V.; *et al*. A global map of rainfed cropland areas (GMRCA) at the end of last millennium using remote sensing, *International Journal of Applied Earth Observation and Geoinformation*, **2009**, 11, 114-129.
5. Thenkabail, P.S.; Biradar, C.M.; Noojipady, P.; Dheeravath, V.; Li, Y.J.; Velpuri, M.; Gumma, M.; Reddy, G.P.O.; Turral, H.; Cai, X.L.; Vithanage, J.; Schull, M.; Dutta, R.. Global irrigated area map (GIAM), derived from remote sensing, for the end of the last millennium. Global Irrigated Area Map (GIAM) for the End of the Last Millennium Derived from Remote Sensing. *International Journal of Remote Sensing*, **2009**, 30, 3679-3733.
6. Portmann, F.T.; Siebert, S.; Doll, P. MIRCA2000 – Global monthly irrigated and rainfed crop areas around the year 2000: a new high-resolution data set for agricultural and hydrological modelling, *Global biogeochemical Cycles*, **2010**, 24, GB1011, DOI:10.1029/2008GB003435.

7. Ramankutty, N.; Evan, A.T.; Monfreda, C.; Foley, J.A. Farming the planet: 1. Geographic distribution of global agricultural lands in the year 2000, *Global biogeochemical cycles*, 2008, 22.
8. Friedl, M.A.; Mciver, D.K.; Hodges, J.C.F.; Zhang, X.Y.; Muchoney, D.; Strahler, A.H.; Woodcock, C.E.; Gopal, S.; Schneider, A.; Cooper, A.; Bacini, A.; Gao, F.; Schaaf, C. Global land cover mapping from MODIS: algorithms and early results. *Remote Sensing of Environment*, 2002, 83, 287–302.
9. Mayaux, P.; Bartholome, E.; Fritz, S.; Belward, A. A new land-cover map of Africa for the year 2000, *Journal of Biogeography*, 2004, 31, 861–877.
10. Defourny, P.; Vancutsem, C.; Pekel, J.F.; Bicheron, P.; Brockmann, C.; Niño, F.; Schouten, L.; Leroy, M. Towards a 300 m global land cover product – the Globcover initiative. In: Braun, M. (ed.) Second Workshop of the EARSeL Special Interest Group on Land Use and Land Cover: Application & Development, Universtätsclub Bonn, Bonn (Germany), **2008**.
11. Fritz, S.; You, L.; Bun, A.; See, L.; McCallum, I.; Schill, C.; Perger, C.; Liu, J.; Hansen, M.; Obersteiner, M.. Cropland for sub-Saharan Africa: A synergistic approach using five land cover data sets, *Geophysical Research Letters*, **2011**, 38.
12. Gregorio, A.D.; Jansen, L. Land Cover Classification System, Classification Concepts and User Manual. *Food and Agriculture Organisation of the United Nations*, **2000**.
13. Fritz, S.; McCallum, I.; Schill, C.; Perger, C.; Grillmayer, R.; Achard, F.; Kraxner, F.; Obersteiner, M.; Geo-Wiki.Org: The use of crowd-sourcing to improve global land cover. *Rem. Sens*., **2009**, 1, 345-354.
14. Vancutsem, C.; Pekel, J-F.; Evrard, C.; Malaisse, F.; Defourny, P.. Mapping and characterizing the vegetation types of the Democratic Republic of Congo using SPOT VEGETATION time series. *International Journal of Applied Earth Observation and Geoinformation*, **2009**, 11, 62–76.
15. Vancutsem, C.; Pekel, J.-F.; Kayitakire, F. Dynamic mapping of cropland areas in sub-saharan Africa using MODIS time series, Proceedings of the 6th International Workshop on the analysis of Multi-temporal remote sensing images, Trento, Italy, July **2011**, IEEE, 25-28, DOI 10.1109/Multi-Temp.2011.6005038, available online: http://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=5983499
16. Fritz, S.; See, L.; McCallum, I.; Schill, C.; Perger, C.; Obersteiner, M. Building a crowd-sourcing tool for the validation of urban extent and gridded population. Computational Science and its Applications – ICCSA 2011. *Lecture Notes in Computer Science*, **2011**, 6783, 39-50.
17. Bontemps S., Defourny P., Van Bogaert E., Herold M., Kooistra L., Kalogirou V. and Arino O. (2012). Producing global land cover maps consistent over time to respond the needs of the climate modelling community. Accepted. *Biogeosciences.*

European Commission
EUR 25570 – Joint Research Centre – Institute for Environment and Sustainability

Title: Harmonizing and combining existing land cover/use datasets for cropland areas monitoring at the African continental scale.

Author(s): Vancutsem Christelle, Marinho Eduardo, Kayitakire Francois, See Linda, Fritz Steffen

Luxembourg: Publications Office of the European Union

2012 – 27 pp. – 21.0 x 29.7 cm

EUR – Scientific and Technical Research series – ISSN 1831-9424 (online)

ISBN 978-92-79-27160-1 (pdf)

doi: 10.2788/61707

Abstract

Mapping cropland areas is of great interest in diverse fields, from crop monitoring to climate change and food security. Recognizing the value of a reliable and harmonized crop mask that entirely covers the African continent, the objectives of this study were to (i) consolidate the best existing land cover/land use datasets, (ii) adopt the Land Cover Classification System (LCCS) for harmonization and (iii) assess the final product. Ten datasets were compared and combined through an expert-based approach to create the derived map of cropland areas at 250m covering the whole of Africa. The resulting cropland mask was compared with two recent cropland extent maps at 1km: one derived from MODIS and one derived from five existing products. The accuracy of the three products was assessed against a validation sample of 3591 pixels of 1km² regularly distributed over Africa and interpreted using high resolution images, which were collected using the agriculture.geo.wiki.org tool. The comparison of the resulting crop mask with existing products shows that it has a greater agreement with the expert validation dataset, in particular for cropland above 30%

LB-NA-25570-EN-N

As the Commission's in-house science service, the Joint Research Centre's mission is to provide EU policies with independent, evidence-based scientific and technical support throughout the whole policy cycle.

Working in close cooperation with policy Directorates-General, the JRC addresses key societal challenges while stimulating innovation through developing new standards, methods and tools, and sharing and transferring its know-how to the Member States and international community.

Key policy areas include: environment and climate change; energy and transport; agriculture and food security; health and consumer protection; information society and digital agenda; safety and security including nuclear; all supported through a cross-cutting and multi-disciplinary approach.

Publications Office

ISBN 978-92-79-27160-1

9 789279 271601