

Article

Assessing Cropland Area in West Africa for Agricultural Yield Analysis

Kaboro Samasse ^{1,2,*}, Niall P. Hanan ³, Gray Tappan ⁴ and Yacouba Diallo ²

¹ Geospatial Sciences Center of Excellence, South Dakota State University, Brookings, SD 57007, USA

² IPR/IFRA, BP 06, Koulikoro Mali; yacdial2005@hotmail.com

³ Plant and Environmental Sciences, New Mexico State University, Las Cruces, NM 88003, USA; nhanan@ad.nmsu.edu

⁴ U.S. Geological Survey Earth Resources Observation and Science Center, Sioux Falls, SD 57198, USA; tappan@usgs.gov

* Correspondence: kaboro.samasse@jacks.sdstate.edu

Received: 8 October 2018; Accepted: 7 November 2018; Published: 10 November 2018



Abstract: Accurate estimates of cultivated area and crop yield are critical to our understanding of agricultural production and food security, particularly for semi-arid regions like the Sahel of West Africa, where crop production is mainly rain-fed and food security is closely correlated with the inter-annual variations in rainfall. Several global and regional land cover products, based on satellite remotely-sensed data, provide estimates of the agricultural land use intensity, but the initial comparisons indicate considerable differences among them, relating to differences in the satellite data quality, classification approaches, and spatial and temporal resolutions. Here, we quantify the accuracy of available cropland products across Sahelian West Africa using an independent, high-resolution, visually interpreted sample dataset that classifies all points across West Africa using a 2-km sample grid (~500,000 points for the study area). We estimate the “quantity” and “allocation” disagreements for the cropland class of eight land cover products in five Western Sahel countries (Burkina Faso, Mali, Mauritania, Niger, and Senegal). The results confirm that coarse spatial resolution (300 m, 500 m, and 1000 m) land cover products have higher disagreements in mapping the fragmented agricultural landscape of the Western Sahel. Earlier products (e.g., GLC2000) are less accurate than recent products (e.g., ESA CCI 2013, MODIS 2013 and GlobCover 2009). We also show that two of the finer spatial resolution maps (GFSAD30, and GlobeLand30) using advanced classification approaches (random forest, decision trees, and pixel-object combined) are currently the best available products for cropland identification. However, none of the eight land cover databases examined is consistent in reaching the targeted 75% accuracy threshold in the five Sahelian countries. The majority of currently available land cover products overestimate cultivated areas by an average of 170% relative to the cropland area in the reference data.

Keywords: land cover & land use; cropland; accuracy assessment; West Africa; Sahel

1. Introduction

Inter-annual variability in crop production associated with climate variability, pests, and diseases is a global concern, particularly for developing countries, where rural communities often lack the capital to help them cope with crop failures and food shortages [1]. Food security is therefore one of the major challenges faced by rural communities in developing countries. In this context, accurate estimates of the cultivated area, as part of crop yield and monitoring programs, are critical to our understanding of agricultural production, food security, and the associated social and economic issues [2]. Remote sensing-based land cover products constitute an important source of information for

analyzing the dynamics of natural and anthropogenic terrestrial ecosystems, particularly for planning food security policies [3]. At national, regional, and global scales, satellite-based systems are necessary, because of their ability to measure large areas, providing timely and consistent data.

Several freely-accessible global land cover products, including agricultural land cover classes, are available at varying spatial resolutions. These products utilize different sources of satellite data and implement different classification approaches, with varying accuracy and spatial resolution [4]. Previous analyses have reported overall and class-specific accuracies at a global scale for some of these land cover products (e.g., [5–8]). However, a more detailed regional assessment, particularly for the West African Sahel, of these global products has so far not been published.

The class-specific accuracy of GLC2000 [8], MODIS collection 5 land cover [7], GlobCover [5], and ESA CCI Land Cover [6] have been reported only at global or continent scales, with important disparities among them, particularly for the cropland classes. At a continent scale, Fritz et al. [9] developed a synergy cropland map (IIASA Cropland) for sub-Saharan Africa, using five global land cover datasets (GLC2000, MODIS Land Cover, GlobCover, MODIS Crop Likelihood, and AfriCover). The combined product has been validated using a Geo-Wiki crowdsourcing application, with reported improvements over the individual datasets for the cropland class [9]. A similar cropland intensity map has been initiated by the Food and Agriculture Organization of the United Nations (FAO) as part of the GLC-SHARE global land cover data. GLC-SHARE aims to provide the global climate modeling community with a baseline product [10].

The overall accuracy or cropland class accuracy may change among regions and continents, because the classification approaches may be more or less successful, and because the availability and quality of the training and reference datasets may vary. For example, at an Africa continent scale, Wei et al. [11] compared the cropland class of five land cover products using Google Earth imagery and the FROM-GLC dataset for validation information. The results showed different accuracies for the different climate zones in Africa. However, the FROM-GLC product itself is found to underestimate the cropland area for African countries [12], and the overall accuracy reported by Wei et al. [11] includes both crop and non-crop classes, without a crop class-specific accuracy assessment. In general, accuracy assessments of land cover products have been done with a less detailed evaluation of the cropland classes at national or regional levels. By using a more accurate reference dataset (the result of manual interpretation of higher spatial resolution images and USGS expert validation; [13]), this paper aims to conduct a detailed performance assessment of the various global land cover datasets so as to accurately map the cultivated area in five Sahelian West Africa countries (Burkina Faso, Mali, Mauritania, Niger, and Senegal). Specifically, we focus on reporting the cropland class user's accuracy (i.e., number correctly identified in a given map class divided by number claimed to be in that map class, related to commission error), the quantity and allocation disagreements based on Pontius and Millones [14], and the good practices of map accuracy assessment suggested in Olofsson et al. [15]. We also report the "area ratio", which is the area of cropland estimated in each global land cover (GLC) product for each country, divided by the area in the reference dataset, as a metric of how well each product defines cropland area.

2. Materials and Methods

Figure 1 describes the different steps for assessing the accuracy of the land cover products for each of the five Sahelian countries, including the eight global datasets included in the analysis, preprocessing, and extraction of the sample points derived from the reference data (details below). For each global land cover product, we created a confusion matrix with error metrics. The final assessment is the comparison of crop areas as identified by the land cover products and the reference data.

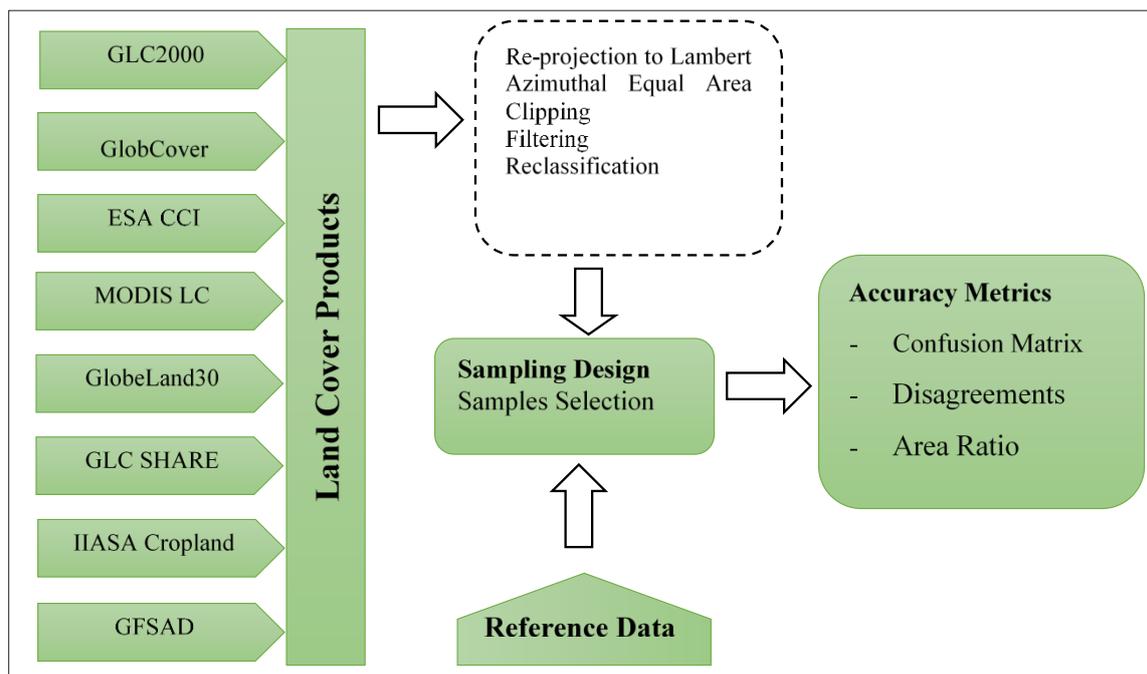


Figure 1. Major steps of the accuracy assessment of different land cover products.

2.1. Reference Data

The West Africa Land Use Dynamics Project (WALUDP) has developed a three-period dataset (1975, 2000, and 2013) to map land use and land cover change across West Africa [13,16,17]. Hundreds of Landsat images with a 30 m spatial resolution (Landsat TM, ETM+, and OLI) and 80 m spatial resolution (Landsat MSS) were sampled at 2 km intervals using the Rapid Land Cover Mapper (RLCM) tool. RLCM was developed by the U.S. Geological Survey (USGS) to facilitate manual image interpretation over large areas and for different periods of time [18]. The sampling consisted of superimposing a grid of dots over the imagery. Each dot of the 2 km by 2 km grid was visually interpreted by experts with local experience in each country. The interpretations were based on Landsat data, with high resolution satellite and aerial photography used to supplement or validate the Landsat classifications. The final dataset provides a classification into one of the 25 land cover types for each centroid of the 2 km grid, with possible land cover classes, including multiple non-agricultural classes, and agricultural classes, including rainfed and irrigated cropland. The approach, based on expert visual interpretation, with specific local knowledge of the environments being classified, is expected to show better results than semi- or fully-automated classifiers, particularly for the cropland land cover type across West Africa [16]. In this study, we used 50% of the 2 km by 2 km data points (selecting data from 2000 or 2013 so as to be closest to the nominal date of the global datasets) as the reference information for assessing the independent land cover products.

Quality control for the reference data was carried out using multiple sources of ancillary data, including thousands of aerial photographs taken by the WALUDP team, high-resolution verification using Google Earth satellite imagery, and field validation in each country, facilitating the systematic verification of land cover assessments [17]. In addition, image interpretation and land cover assessments carried out by national experts were reviewed and revised during regular collaborative workshops in West Africa, in order to ensure consistent practice between country teams and USGS partners.

2.2. Land Cover Products

The land cover products assessed in this analysis are shown in Table 1. In this paper, we focus on the accuracy assessment of crop classes rather than on assessing the performance of non-crop land cover classes.

Table 1. Characteristics of the different Land Cover Products Crop classes. The numbers in brackets (e.g., in Cropland (4)), are the class numbers, as reported by the producers.

	Product	Crop Classes	Mixed Crop Classes	Year	Resolution (m)
Group I	GLC2000	Croplands (>50%) (18); tree crops (21); irrigated croplands (20)	Mosaic Forest/Croplands (7); Croplands with open woody vegetation (19)	2000	1000
	GlobCover	Post-flooding or irrigated croplands (11); rainfed croplands (14)	Mosaic cropland (50–70%)/vegetation (grassland, shrubland, and forest) (20–50%) (20); mosaic vegetation (grassland, shrubland, and forest) (50–70%)/cropland (20–50%) (30)	2005;2009	300
	ESA CCI LC series	Cropland rainfed (10); cropland irrigated (20)	Mosaic cropland (30)	1992–2015	300
	ESA CCI 20 m	Cropland (4)		2016	20
	GlobeLand30	Cultivated land (10)		2010	30
	GFSAD 30 m Crop Extent	Croplands (2)		2015	30
	MODIS Land Cover (MCD12Q1)	Croplands (12)	Cropland/natural vegetation mosaic (14)	2001–2013	500
Group II	IIASA IFPRI Cropland	Crop intensity greater than 50%	Crop intensity less than 50%	2005	1000
	GLC SHARE	Crop intensity greater than 50%	Crop intensity less than 50%	2014	1000

2.2.1. GLC2000

The global land cover database for the year 2000 (GLC2000) has been coordinated by the European Commission's Joint Research Center (JRC) in collaboration with 30 research partners (<http://forobs.jrc.ec.europa.eu/products/glc2000/products.php>). The global product is generated based on regional products defined by national and regional experts across the world, using the SPOT-Vegetation data [19]. In total, the global product has 22 land cover types at a 1 km spatial resolution. The land cover classes are compatible with the United Nations Land Cover Classification System (UN-LCCS) [19].

2.2.2. GlobCover

GlobCover is an initiative of the European Space Agency (ESA) to produce a global land cover product at a finer spatial resolution than the GLC2000 (http://due.esrin.esa.int/page_globcover.php). The first product (GlobCover V2.2) is centered on years 2005–2006 [20] and the second (GlobCover V2.3) is on year 2009 [21]. The ENVISAT MERIS sensor data at a 300 m spatial resolution is the main input data for the GlobCover product. Given the improvements in the underlying data and the classification technique in V2.3 [21], we have focussed more on the performance of the more recent product.

2.2.3. ESA LC CCI

The European Space Agency Land Cover of the Climate Change Initiative project (ESA LC CCI) is a set of multi-sensor global land cover databases (<https://www.esa-landcover-cci.org/>). Recently, the ESA has released a time series of land cover maps, from 1992 to 2015, at a 300 m pixel size [22]. A combination of ENVISAT MERIS, SPOT-Vegetation, and ASAR instruments are used to develop a consistent global land cover product, as a European contribution to the Essential Climate Variables (ECV) list required by the United Nations Framework Convention on Climate Change (UNFCCC). The availability of Sentinel-2 time-series data has also contributed to the development of a prototype for the existing finest spatial resolution (20 m) land cover database of Africa for the year 2016. Both the 300 m (CCI) and 20 m (CCI-20) land cover maps are assessed in this study.

2.2.4. GlobeLand30

A 30-m global land cover product was recently developed by Chinese National Geomatics Center for two periods, 2000 and 2010 (<http://www.globeland30.org/GLC30Download/index.aspx>). An ensemble of classifiers based on the integration of pixel- and object-based land cover classification were used, with expert knowledge for better handling of spectral confusion and diversity of complex landscapes across the globe [23]. Thousands of Landsat images, together with the Chinese HJ-1 satellite images, served in the development of GlobeLand30 maps.

2.2.5. MODIS Land Cover

The Moderate Resolution Imaging Spectroradiometer (MODIS) collection 5.1 Land Cover product (MCD12Q1 at 500 m) was also assessed in this study (https://lpdaac.usgs.gov/dataset_discovery/modis/modis_products_table/mcd12q1). MCD12Q1 is a yearly global land cover dataset, covering years 2001 to 2013, derived from both Terra and Aqua observations, using five global land cover classification systems [7]. We used the International Geosphere Biosphere Programme (IGBP) classification scheme with 17 land cover classes developed using an ensemble of decision trees based on training data and ancillary data layers, with noise reduction and quality assessments (as described by the authors of [7]).

2.2.6. Global Land Cover SHARE

The Global Land Cover SHARE (GLC-SHARE; <http://www.fao.org/geonetwork/srv/en/main.home>) was developed by the Food and Agriculture Organization of the United Nations (FAO) as a merger of national, regional, and global databases, with a high and medium spatial resolution (30 m or less) and ~66% global coverage. In the absence of high resolution national and regional data, coarser-scale land cover estimates were used. The multi-temporal and multi-source data were then harmonized and standardized using a data fusion approach based on the Land Cover Classification System (LCCS) and the Land Cover Meta Language (LCML) elements. The final product has a spatial resolution of ~1 km, with 11 aggregated land cover types represented with 0 to 100% of the area covered in each pixel [10].

2.2.7. IIASA IFPRI Cropland Map

By adopting a hybrid data integration approach, the authors of [24] developed a global cropland percentage map at a 1 km spatial resolution, referred to as the International Institute for Applied Systems Analysis-International Food Policy Research Institute (IIASA IFPRI) cropland product (<http://www.iiasa.ac.at/web/home/about/news/150116-Cropland-Maps.html>). Several global land cover products combined, with regional and national datasets used as the inputs to create this cropland intensity map, which were validated using FAO agricultural statistics data.

2.2.8. GFSAD Crop Extent for Africa

This product was developed by the NASA Global Food Security Support Analysis Data (GFSAD) combining Landsat 8 and Sentinel-2 data. A combination of pixel- and object-based classification approaches has been used to develop the dataset. Specifically, random forest, support vector machine, and the recursive hierarchical segmentation (RHSeg) algorithms were used to improve the performance of the data classification for cropland mapping. The product has a 30 m spatial resolution with 2015 as the nominal assessment year [25].

2.3. Accuracy Assessment

2.3.1. Sampling Design

The global land cover datasets were separated into two groups so as to differentiate those with multiple land cover classes from those with only crop intensity information (Table 1). The land cover

classes for class-based products were redefined as “cropland”, “mixed cropland”, and “non-cropland”, based on the class descriptions associated with each product. For the second group, we reclassified the crop intensity according to the percentage cropland, with 50–100% of the crops defined as “cropland”, and less than 50% defined as “mixed crop” (Table 1). We then sampled the reclassified (crop, mixed crop, and non-crop) land cover datasets using the coordinates of the reference data (described above), providing systematic (regular) reference land cover assessments every 4 km across the entire region (257,724 sample points; Figure 2). That is, for each reclassified land cover, at each of these 257,724 points, we extracted the pixel value (equivalent to the category or land cover class). An error matrix is then created using the extracted pixel values.

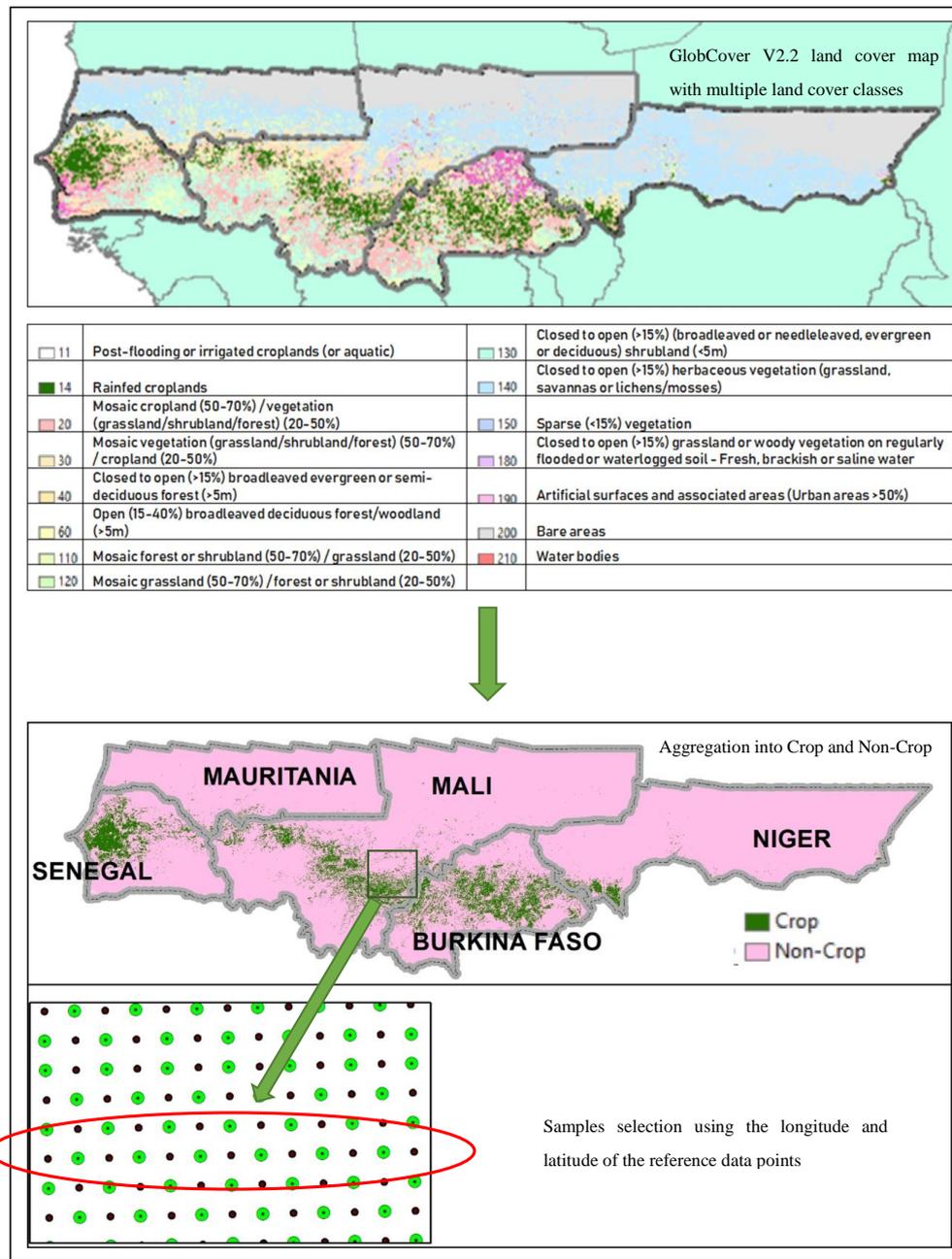


Figure 2. Example of the GlobCover V2.2 land cover map, aggregated into crop and non-crop classes, and a sample selection for error assessments. The first step was to aggregate the land cover classes with respect to the presence of agricultural activity. The second step extracted land cover data based on latitude and longitude of the reference data set (sampling every second location; green circles).

2.3.2. Metrics of Accuracy

Pontius and Millones [14] suggested a method to assess the accuracy of classified maps derived from the remote sensed data. The method is based on two simple concepts, quantity disagreement (Q) and allocation disagreement (A). Q is defined as the difference between the reference classes and the map classes, which is due to the mismatch in the proportions of the different classes. For our study, the Q can be considered to be a measure of error in “how much cropland” there is. A is the difference between the reference classes and the map classes, which is due to a mismatch in the spatial location of the categories. For our study, A translates as error in “where the cropland is”. The total disagreement is the sum of Q and A [14]. The calculation is based on a stratified sampling method. Each land cover class (in our case, crop, mixed crop, and non-crop) is considered as a stratum with a number of pixels, N. The sample confusion matrix (Table 2) is created by extracting the pixel values corresponding to the reference data points within each country. From this sample table, we then estimated the population confusion matrix (Table 3) for a random or systematic stratified sampling, using Equation (1).

Table 2. Sample confusion matrix for two aggregated land cover types (i.e., GLC group 2).

	Ref.	1. Crop/Mixed Crop	2. Non-Crop
GLC			
1. Crop/Mixed Crop		n_{11}	n_{12}
2. Non-Crop		n_{21}	n_{22}

Table 3. Population confusion matrix for two aggregated land cover types (i.e., global land cover (GLC) group 2).

	Ref.	1. Crop/Mixed Crop	2. Non-Crop
GLC			
1. Crop/Mixed Crop		p_{11}	p_{12}
2. Non-Crop		p_{21}	p_{22}

The p_{ij} represents the estimate of the area proportion of the population that has class i of the global land cover product and class j for the reference data [15].

$$p_{ij} = \left(\frac{n_{ij}}{\sum_{j=1}^J n_{ij}} \right) \left(\frac{N_i}{\sum_{i=1}^J N_i} \right) \quad (1)$$

where

$j = 1 \dots J$ is the number of classes in the reference data,

$i = 1 \dots J$ is the number of classes in the global land cover product,

$\sum_{j=1}^J n_{ij}$ is the sample total for class i ,

N_i is the population total for class i

Summary of the confusion matrix

The quantity disagreement (Q) and allocation disagreement (A) are estimated using the population table [14], expressed by Equations (2) to (5).

$$q_k = \left| \left(\sum_{i=1}^J p_{ik} \right) - \left(\sum_{j=1}^J p_{kj} \right) \right| \quad (2)$$

q_k : quantity disagreement for class k

$$Q = \frac{\sum_{k=1}^J q_k}{2} \quad (3)$$

$$a_k = 2 \times \min \left[\left| \left(\sum_{i=1}^J p_{ik} \right) - p_{kk} \right|, \left| \left(\sum_{j=1}^J p_{kj} \right) - p_{kk} \right| \right] \quad (4)$$

a_k : allocation disagreement for class k

$$A = \frac{\sum_{k=1}^J a_k}{2} \quad (5)$$

The overall accuracy (OA) or proportion correct is estimated using Equation (6), and the user's accuracy (UA) for a specified class is given by Equation (7).

$$OA = \sum_{i=1}^J p_{ii} \quad (6)$$

$$UA_k = \frac{p_{kk}}{\sum_{i=1}^J p_{ki}} \quad (7)$$

3. Results

3.1. All Crop (Crop and Mixed Crop) Quantity and Allocation Disagreements

Here, we focused on the overall accuracy of the cropland designations in the global land cover products, defining "all crop" to be the sum of the "crop" and "mixed crop" classes. Figure 3 shows the results obtained in terms of the disagreements in the crop class, organized by country. The proportions are expressed in terms of quantity disagreement and allocation disagreement. In the majority of cases, the most important part in the total disagreement is as a result of the quantity disagreement. This is particularly the case in Mauritania, where the disagreement due to allocation is negligible compared with the quantity disagreement, resulting in a large overestimation in the total number of pixels identified as crop in the land cover products, relative to the reference data. The maximum total disagreement occurs with the GLC2000 product in Burkina Faso (50%), Mali (28%), and Senegal (66%). The maximum disagreement in Niger is observed with GLOBCOVER2009 (26%), and in Mauritania with ESACCI20.2016 (30%). The 20 m spatial resolution land cover product of ESA CCI (for 2016) is significantly less accurate among the databases for cropland mapping in Mauritania, as compared to the other countries in the study area. Qualitatively and quantitatively, the most accurate crop predictions are from GlobeLand30 and GFSAD30. These products both have a 30 m spatial resolution, using Landsat images as the main inputs for the classification. This finding regarding GlobeLand30 is in line with previous studies at the Africa continent level [11].

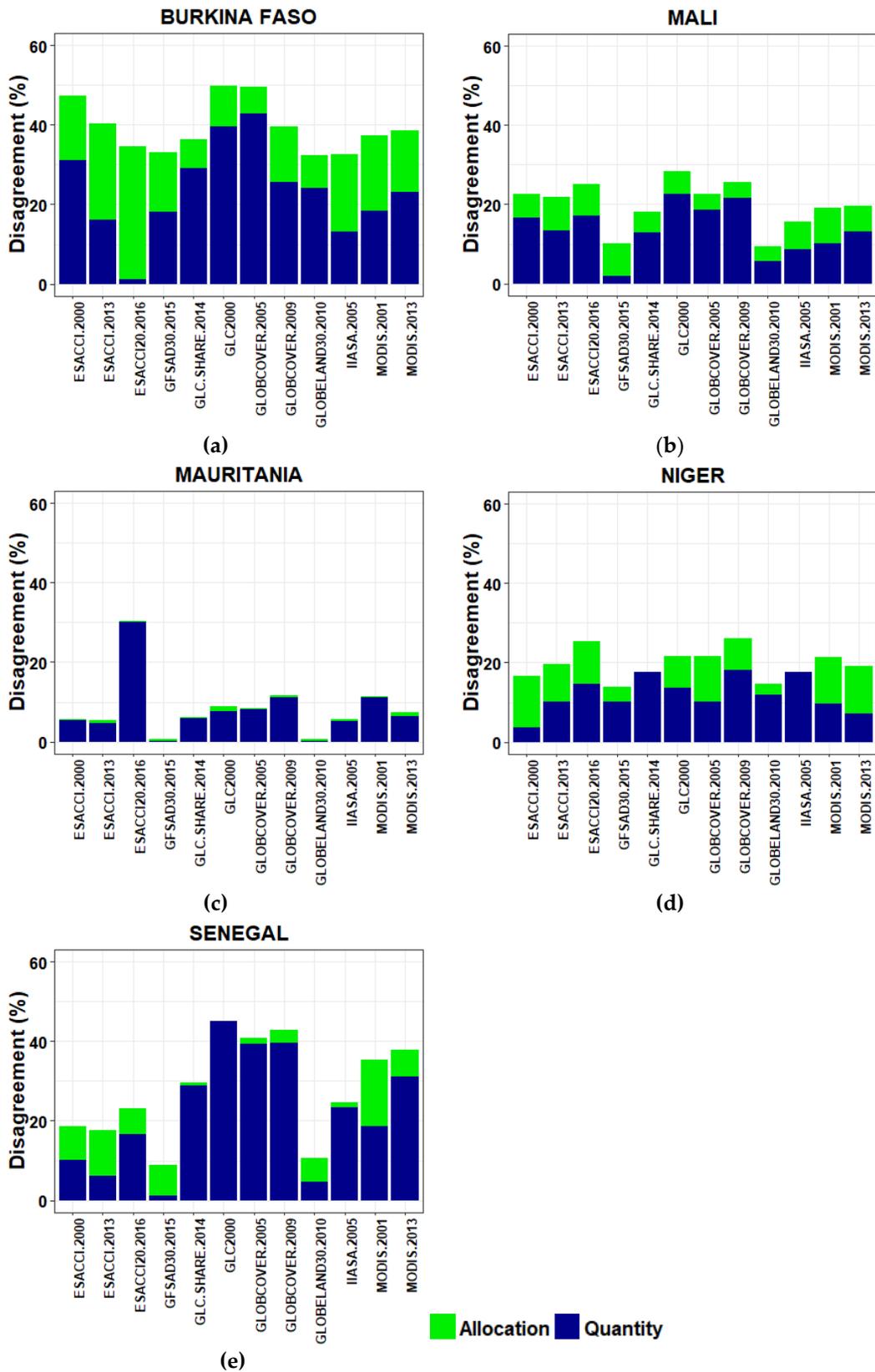


Figure 3. All cropland (sum of crop and mixed-crop classes) quantity disagreement (quantity = how much cropland) and allocation disagreement (allocation = where the cropland is) in five Sahelian West African countries: Burkina Faso (a), Mali (b), Mauritania (c), Niger (d) and Senegal (e). Twelve independent land cover assessments (eight “products”, some with multiple years) are used.

3.2. Cropland User's Accuracy

3.2.1. All Crop (Crop and Mixed Crop) User's Accuracy

Metrics like the user's and producer's accuracy are often reported to bring additional class-specific information to remote sensing-based classifications. In this study, we are interested in assessing how well land cover products classify crop classes (i.e., the user's accuracy). Based on Figure 4, it is evident that GlobeLand30 and GFSAD30 present the best accuracy for cropland mapping. GLC2000 has the greatest misclassification proportion. Mauritania is the country among the five involved that are in this research where the land cover products did not perform well in locating crop areas. Half of the land cover products completely failed (accuracy ~0%) to identify cropland correctly. Maybe because of the particularly small size of farms in this country. Mauritania also has the least agricultural land area in West Africa, mostly localized along the Senegal River. It is also worth mentioning that the 20 m land cover product of Africa (ESACC20) has a similar or worse accuracy than the 300 m ESA land cover products.

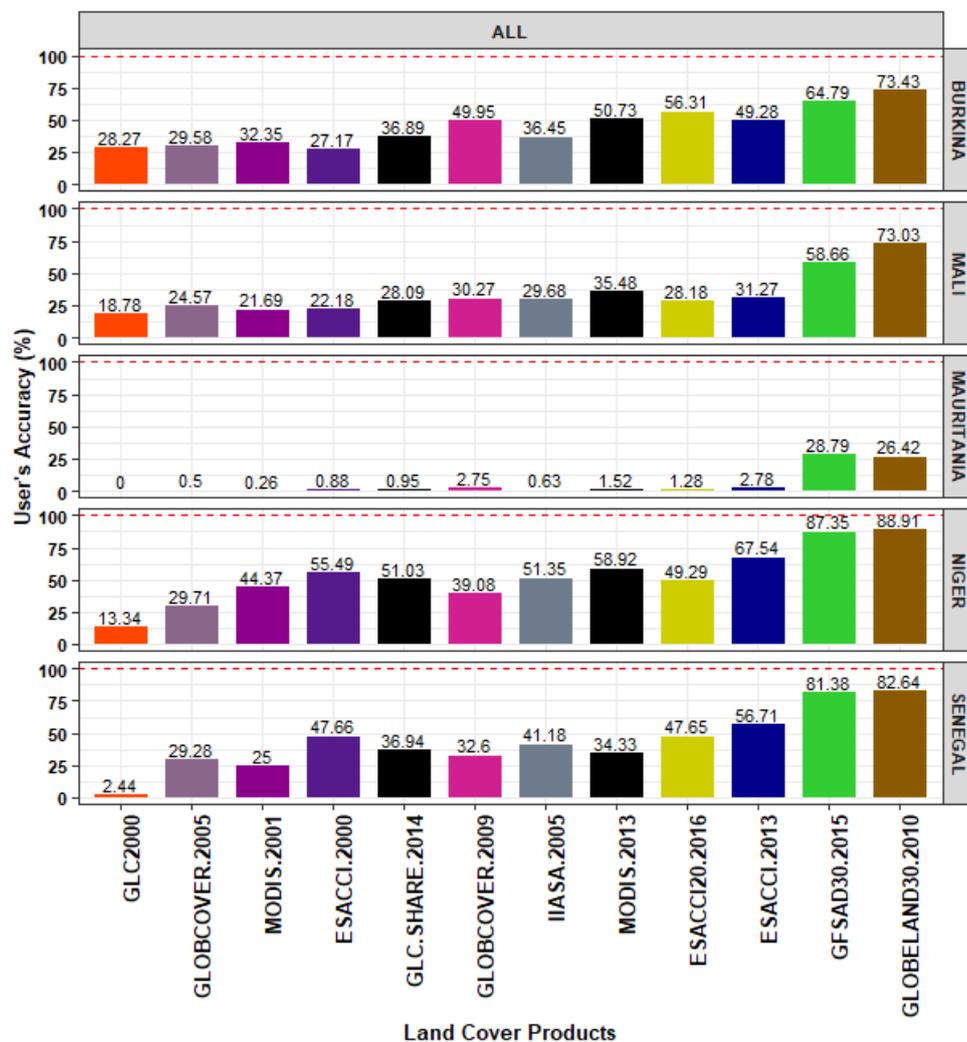


Figure 4. User's accuracy of the "all crop" class by country and land cover product.

On average, for all countries and land cover products, the user accuracy is 35.97% (Figure 5a). Figure 5b shows that the low accuracy in Mauritania reduces the overall average user’s accuracy. Averaging the user’s accuracy without Mauritania increases the metric to 43.57% (Figure 5c). Cropland mapping accuracy among land cover products shows that GFSAD30 and GlobeLand30 present the best accuracies (with and without the inclusion of Mauritania). In the majority of cases, land cover products have less than 50% of the cropland pixels identified, based on the reference information. Most of the more recent land cover datasets (those developed after 2010, except for the FAO GLC SHARE) present accuracies above the average, indicating gradual improvements related to the availability of more recent remote sensing datasets, calibration data, and analytics. Those datasets are the 2013 ESACCI with a 300 m pixel size (51.2%), ESACCI prototype with a 20 m for 2016 (45.4%), GFSAD with a 30 m for 2015 (73.1%), GlobeLand30 with a 30 m for 2010 (79.5%), and MODIS with a 500 m for 2013 (44.9%).

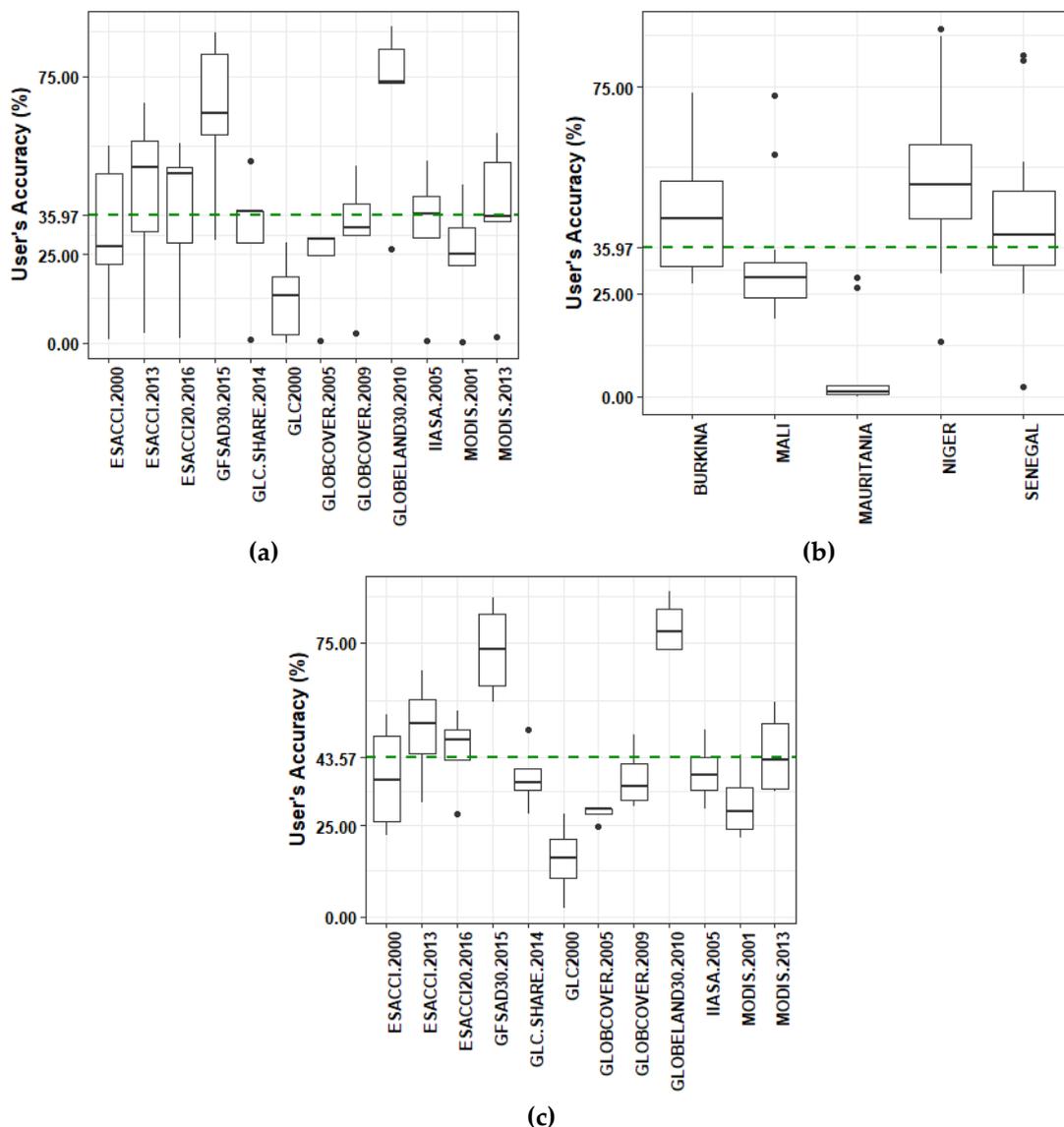
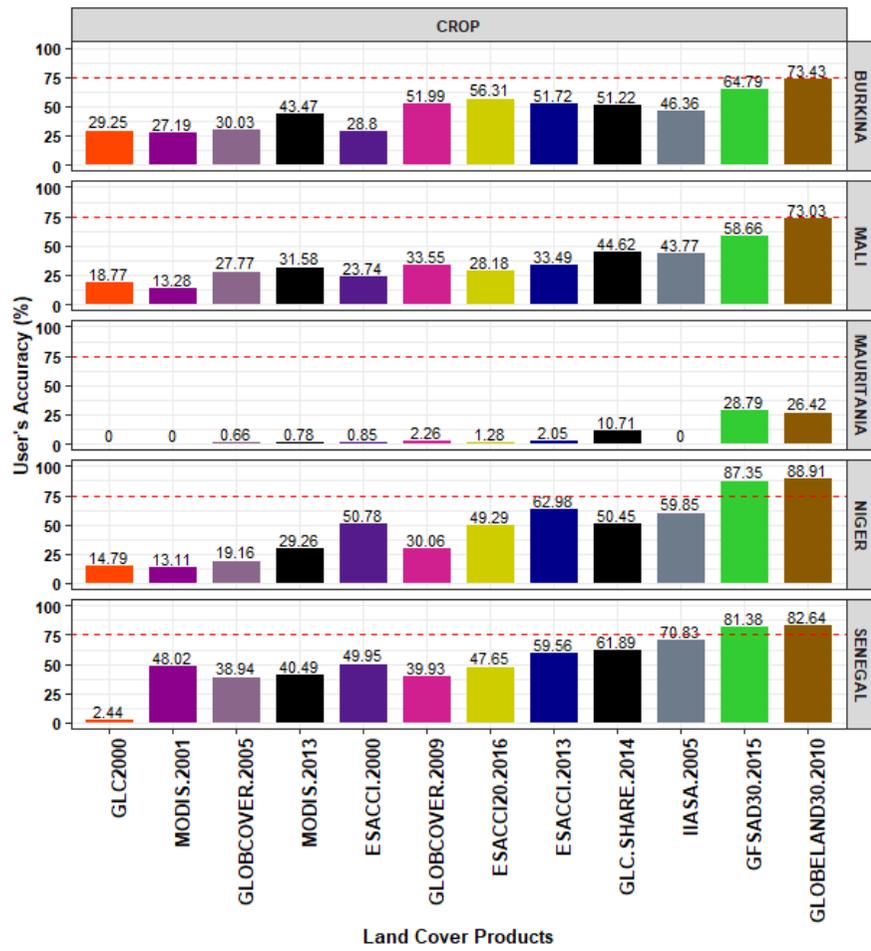


Figure 5. User’s accuracy of crop class for all countries (a), all GLC products (b), and all GLC products, with Mauritania removed (c).

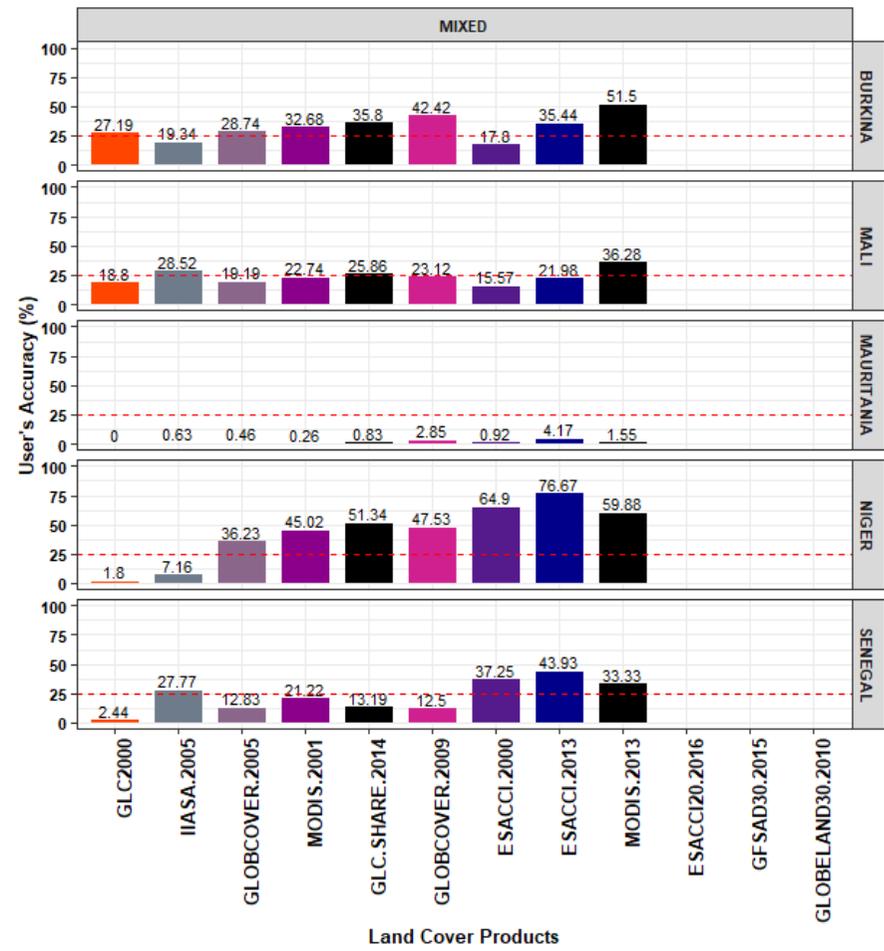
3.2.2. User’s Accuracy for Separate Crop and Mixed Crop Classes

The two land cover classes (“crop” and “mixed-crop”; Table 1) are analyzed separately in this section. Land cover products with a unique agricultural class, like ESA CCI at 20 m, GFSAD30,

and GlobeLand30, are analyzed in the category of crop class without a mixed crop (Table 1). The products with the crop intensity in a percentage, like GLC SHARE and the IIASA crop intensity, are analyzed with separate crop and mixed crop classes. A striking illustration of user's accuracy through the two above-mentioned crop classes can be seen in Figure 6. Each of the five countries (Burkina Faso, Mali, Mauritania, Niger, and Senegal) are presented by crop class and land cover product. The two thresholds (see Figure 5), one at 75% and another at 25%, are taken as the targeted values for crop and mixed crop classes, respectively. The assumptions being that the crop class pixels with a 50–100% crop cover are identified as cropland (hence an average value of 75%), while pixels with a 0–50% crop cover are identified as the mixed crop (hence an average value of 25%). For the crop and mixed crop categories, land cover products with explicit details on the fraction of cropland in pixels are considered in the analysis. Those products are ESA CCI, GLC2000, GlobCover, GLC SHARE, IIASA Cropland, and MODIS land cover datasets (Table 1). None of the land cover products reach the accuracy of 75% for crop class for all of the five countries. GFSAD30 and GlobeLand30 have the most accurately mapped cropland in this category, with accuracies exceeding 80% in some countries. For example, in Niger and Senegal, these accuracies are 87.35% and 81.38%, respectively, for GFSAD30, and 88.91% and 82.64%, respectively, for GlobeLand30. They are followed by the ESA CCI land cover of year 2013, GLC SHARE, and IIASA crop intensity map. GlobCover and MODIS have a similar crop accuracy in Burkina Faso and Senegal. GLC2000 remains the least accurate in this category of crop class in terms of user's accuracy. Contrary to the crop class, the mixed crop class has been relatively well identified, as the majority of land cover products have more than 25% of mixed crop user's accuracy. On average, the MODIS and ESA CCI land cover products for 2013 have the best accuracies in locating pixels with mixed cropland. They both have a coarse resolution, of 500 m and 300 m, respectively. At the country level, Mauritania is shown as being exceptionally poorly mapped in terms of cropland for both of the two categories of the crop classes.



(a)



(b)

Figure 6. User's accuracy for crop (a) and mixed crop (b) classes. ESACCI 20 m, GFSAD30, and GlobeLand30 do not have mixed crop class (b).

3.3. Cropland Area Assessment

3.3.1. All Crop Commission and Omission Disagreements

The reported omission disagreement and commission disagreement in Figure 7 correspond to false negative and the false positive outcomes [or cases], respectively. They are equivalent to the off-diagonal terms of the crop/mixed crop class p_{12} (commission) and p_{21} (omission), as defined in Table 3. The agreement value is the diagonal term p_{11} for the crop/mixed crop class. A higher value of commission disagreement compared to omission disagreement is similar to an overestimation cropland. On the other hand, if the omission disagreement is greater than the commission disagreement, the cropland is underestimated by the land cover product. Any difference between these two metrics of disagreement means a non-zero quantity disagreement from the crop class. Based on this analysis, the GLC2000, GlobCover 2005, GLC SHARE, ESACCCI 2000, and MODIS 2013 products visibly present the most important difference between omission and commission disagreements. Therefore, they are expected to have a greater overestimation of crop area. GlobeLand30 and GFSAD30 have omission disagreements greater than commission disagreements, leading to an underestimation of crop area. This underestimation is expected to be more important in the GlobeLand30 compared to GFSAD30.

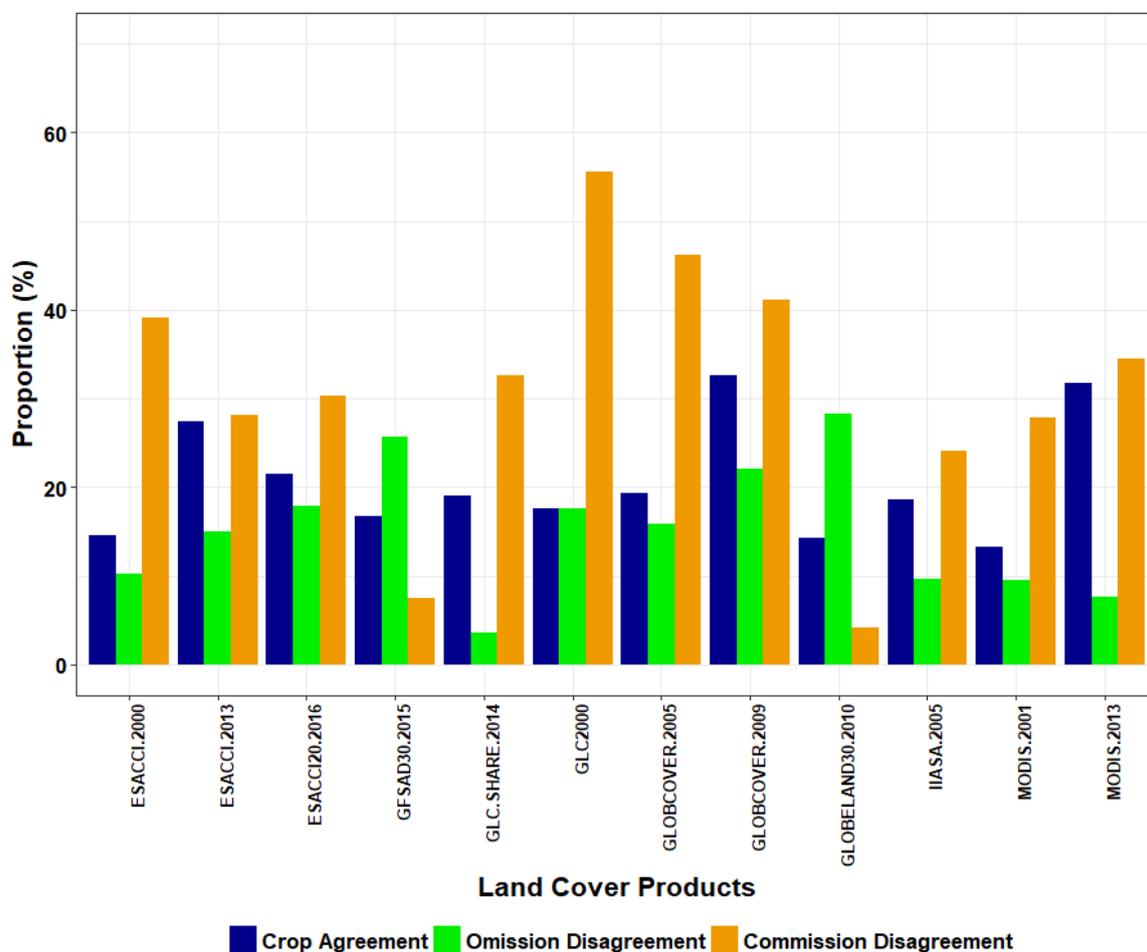


Figure 7. Crop proportion correct with the relative omission and commission misclassifications. Higher commission to omission disagreement = overestimation of cropland, while higher omission to commission disagreement = underestimation of cropland.

3.3.2. Cropland Area Ratio (GLC/Reference Map)

Figure 8 illustrates the general trends of the GLC products cropland area estimation in the study region. These trends are mainly an overestimation of the crop area. At the country level, land cover

products have overestimated the cropland area by a factor greater than 20 in Mauritania, while this factor has an average of less than 2 in other countries (Figure 8a). For this reason, the aggregation of area ratios by land cover product (Figure 8b) is carried out without Mauritania, so as to reduce bias in the results.

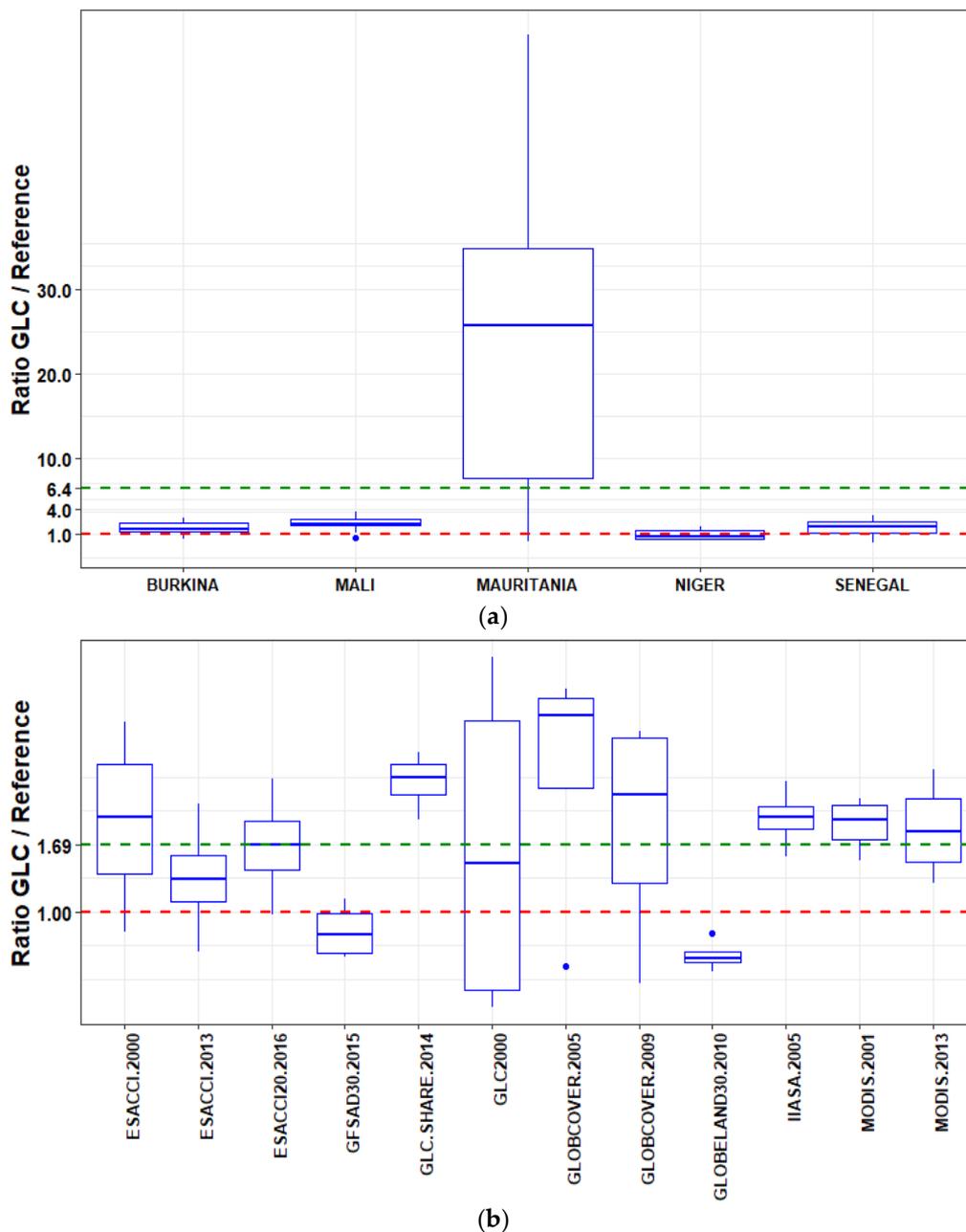


Figure 8. Ratio of mapped and reference crop areas. Aggregated by country (a), and by land cover products (b) with the ideal situation (ratio = 1) in red and the observed average ratio (1.69) in green.

The generalized overestimation of crop areas shown in Figure 8b is in line with what has been previously discussed, relative to the difference between omission and commission disagreements. A ratio greater than 1 indicates a tendency, across most of the GLC products, to an overestimation of the cropland area. On average, the GLC datasets overestimate the cropland area by about 69%, although two GLCs (GFSAD30 with a ratio of 0.8 and GlobeLand30's with a ratio of 0.6) underestimate the cropland area. Even if the average ratio of GLC2000 is less than the overall average, this product presents the most substantial variations in crop areas estimation, followed by GlobCover.

4. Discussion

Recently developed land cover datasets (after 2010) seem to have a better accuracy in cropland mapping in the Sahel region (Figures 4 and 6). This is due not only to the availability of newer sensors with higher spatial, spectral, and radiometric resolutions (e.g., Landsat 8: <https://landsat.usgs.gov/landsat-8>, and Sentinel 2: <https://sentinel.esa.int/web/sentinel/home>), but also to the progress made in the implementation of new approaches of satellite image processing, particularly in machine learning techniques (Support Vector Machine, Decision trees, Random forest, Segmentation algorithm). GFSAD30 and GlobeLand30, which have shown better cropland estimation in Western Sahel, compared to the other land cover products, were developed using random forest and pixel-object-based (i.e., an optimization of the pixel-based and object-based methods) classification algorithms, respectively. These new algorithms are becoming popular within the remote sensing community, because of their abilities to accurately classify land cover [26,27]. However, at the country and regional scales of this West Africa analysis, using high quality and high density reference information, we found that both GFSAD30 and GlobeLand30 have a cropland class user's accuracy below that reported by the producers; which is in conformity with previous findings in assessing the GlobeLand30 dataset at a country level (Kenya) reported by the authors of [28].

In general, the land cover maps with multiple classes based on coarse spatial resolution (i.e., 300 m or greater) satellite images did not perform well in identifying crop areas in the region of study. At this coarse spatial resolution, it is common to find cropland mixed with fallow and other cover types in the fragmented Sahelian agricultural landscape, adding another level of complexity in land cover classification [29]. The best user's accuracy for the mixed cropland class is shown with the European climate change initiative product of the year 2013 at 300 m (ESACCI 2013). This is in large part because the moderate spatial resolution products (i.e., 30 m and 20 m) do not have a mixed class for cropland. GLC2000, developed around the year 2000, with its 1 km pixel size, has the highest disagreement values and the most variable crop area ratio across the five countries. Some improvements have been made with the ESA CCI land cover series, the GlobCover at 300 m spatial resolution, and MODIS land cover at 500 m. These products have multi-year data (ranging from 1992 to 2015 for ESACCI, 2004 to 2009 for GlobCover, and 2001 to 2013 for MODIS). The results of this study suggest that multi-year products are improving through time, perhaps because of better classification approaches (e.g., handling mixture of cover types in pixels), as data are coming from the same sensors. However, we found a general overestimation of crop areas in this category of land cover product. The specific case of ESA CCI's overestimation of cropland area has been concluded by Laso Bayas et al. [12] across the Africa continent in previous studies.

Attempts to improve land cover products, particularly for mapping the crop extent, have led to the development of hybrid products with cropland intensity (from 0 to 100%), created by fusing various data sources with different existing land cover datasets. Data fusion is well known approach in remote sensing. Its' goal is to obtain a higher reliability by using multi sources data [30]. These synergistic or hybrid products are normally at coarser spatial resolution (~1 km) and include GLC SHARE and IIASA Crop intensity products assessed in this study, which both overestimate cropland in West Africa with crop area ratios above the overall average (i.e., 1.69) (Figure 8b).

Our results confirm the general observation that coarse pixel size is not suitable for mapping the fragmented cropland landscapes in the West African Sahel. However, expectations for the improved identification of small agricultural fields in West Africa are generally not met with higher resolution products, including the recently released 20 m ESA CCI land cover product based on Sentinel 2 images. Indeed, ESA CCI presents worse agreement in mapping cropland than the 30-m based products like GFSAD30 and GlobeLand30. Even if the year of production could be a factor in the disagreement between those three products (2016 for ESACCI 20 m, 2010 for GlobeLand30, and 2015 for GFSAD30), this raises questions about the approach and training data used to develop the 20 m ESA product for Africa. The spectral and structural similarities of cropland with the surrounding natural vegetation

(shrubs, grassland, savanna, and fallow) could also lead to difficulties in correctly mapping crop areas in the region, despite improvements in the spatial resolution (from 1000 m to 20 m).

5. Conclusions

In total, eight land cover products were assessed in this study, with a focus on cropland classes across five Western Sahel countries (Burkina Faso, Mali, Mauritania, Niger, and Senegal). These products have been developed for diverse purposes and present different characteristics in terms of input data, algorithm of classification, and consistency.

In general, a low user's accuracy of cropland class and high crop area ratios (overestimation) are observed with coarse spatial resolution (i.e., 300 m or greater) land cover products. However, these products seem to map accurately the mixed cropland class, as the majority of land cover products in this category have more than 25% of the mixed crop user's accuracy. ESACCI 2013 for example reaches a user's accuracy of 76.67% in Niger.

Progress in computational power, combined with the availability of new sensors and optimized algorithms, have led to the development of improved land cover datasets. These datasets at 30 m (GFSAD30 and GlobeLand30) or 20 m (ESACCI 2016) are created sometimes by fusing more than one source of data (e.g., Landsat 8 and Sentinel2). However, while GFSAD30 and GlobeLand30 have shown better accuracy and improvement in the crop area ratio compared to the coarser pixel size products, similar expectations are not met with the 20 m ESA CCI land cover.

Overall, among the studied land cover products, GFSAD30 and GlobeLand30 present better accuracy in identifying crop areas. They have, in the Sahel, an average cropland class accuracy of 68.89% and 64.19% for GlobeLand30 and GFSAD30, respectively, approaching the target accuracy of 75%, although both tend to underestimate crop areas. Given the importance of agriculture for food security and livelihoods in West Africa, the development of remote sensing-based approaches to monitoring agricultural yields is of critical importance. The accurate geolocation and area quantification of the croplands is a necessary first step. Our results suggest a considerable variability in the accuracy of the cropland assessments available in the GLC products. However, gradual improvements associated with newer sensors and higher spatial resolution, coupled with innovations in analytical approaches, have led to increases in the overall accuracy as well as the decreasing quantity and allocation errors. New training and validation datasets, derived using expert local knowledge, facilitate the error assessment of the global land cover products, and open the door to locally optimized agricultural land use and land cover assessments.

Author Contributions: K.S. and N.P.H. conceived the analysis. K.S. carried out the analysis and wrote the manuscript. G.T. contributed the underlying datasets. All authors contributed to writing and editing the manuscript.

Funding: K.S. was supported by the United States Agency for International Development (USAID) via funding to the Michigan State University Borlaug Higher Education for Agricultural Research and Development (BHEARD) Graduate Research Fellowship Program.

Acknowledgments: This material is based upon work supported by the United States Agency for International Development, as part of the Feed the Future initiative, under the CGIAR Fund, award number BFS-G-11-00002, and the predecessor fund the Food Security and Crisis Mitigation II grant, award number EEM-G-00-04-00013.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Conceição, P.; Fuentes-Nieva, R.; Horn-Phathanothai, L.; Ngororano, A. Food security and human development in Africa: Strategic considerations and directions for further research. *Afr. Dev. Rev.* **2011**, *23*, 237–246. [[CrossRef](#)]
2. Zhong, L.; Gong, P.; Biging, G.S. Efficient corn and soybean mapping with temporal extendability: A multi-year experiment using Landsat imagery. *Remote Sens. Environ.* **2014**, *140*, 1–13. [[CrossRef](#)]

3. Hüttich, C.; Herold, M.; Wegmann, M.; Cord, A.; Strohbach, B.; Schullius, C.; Dech, S. Assessing effects of temporal compositing and varying observation periods for large-area land-cover mapping in semi-arid ecosystems: Implications for global monitoring. *Remote Sens. Environ.* **2011**, *115*, 2445–2459. [[CrossRef](#)]
4. Herold, M.; Mayaux, P.; Woodcock, C.; Baccini, A.; Schullius, C. Some challenges in global land cover mapping: An assessment of agreement and accuracy in existing 1 km datasets. *Remote Sens. Environ.* **2008**, *112*, 2538–2556. [[CrossRef](#)]
5. Arino, O.; Bicheron, P.; Achard, F.; Latham, J.; Witt, R.; Weber, J.-L. The most detailed portrait of Earth. *Eur. Space Agency* **2008**, *136*, 25–31.
6. ESA-CCI. Internal Release of Global Land Cover Map with Improved Accuracy over the Existing State of the Art (75.6%). Available online: <https://www.esa-landcover-cci.org/?q=node/148> (accessed on 11 January 2013).
7. Friedl, M.A.; Sulla-Menashe, D.; Tan, B.; Schneider, A.; Ramankutty, N.; Sibley, A.; Huang, X. MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote Sens. Environ.* **2010**, *114*, 168–182. [[CrossRef](#)]
8. Mayaux, P.; Eva, H.; Gallego, J.; Strahler, A.H.; Herold, M.; Agrawal, S.; Naumov, S.; De Miranda, E.E.; Di Bella, C.M.; Ordoyne, C. Validation of the global land cover 2000 map. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 1728–1739. [[CrossRef](#)]
9. 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. *Geophys. Res. Lett.* **2011**, *38*. [[CrossRef](#)]
10. Latham, J.; Cumani, R.; Rosati, I.; Bloise, M. *Global Land Cover Share (GLC-SHARE) Database Beta-Release Version 1.0-2014*; FAO: Rome, Italy, 2014.
11. Wei, Y.; Lu, M.; Wu, W. A Comparative Analysis of Five Cropland Datasets in Africa. *ISPRS Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2018**, *XLII-3*, 1863–1870. [[CrossRef](#)]
12. Laso Bayas, J.C.; See, L.; Perger, C.; Justice, C.; Nakalembe, C.; Dempewolf, J.; Fritz, S. Validation of automatically generated global and regional cropland data sets: The case of Tanzania. *Remote Sens.* **2017**, *9*, 815. [[CrossRef](#)]
13. Tappan, G.G.; Cushing, W.M.; Cotillon, S.E.; Mathis, M.L.; Hutchinson, J.A.; Dalsted, K. *West Africa Land Use Land Cover Time Series*; U.S. Geological Survey: Sioux Falls, SD, USA, 2016.
14. Pontius, R.G., Jr.; Millones, M. Death to Kappa: Birth of quantity disagreement and allocation disagreement for accuracy assessment. *Int. J. Remote Sens.* **2011**, *32*, 4407–4429. [[CrossRef](#)]
15. Olofsson, P.; Foody, G.M.; Herold, M.; Stehman, S.V.; Woodcock, C.E.; Wulder, M.A. Good practices for estimating area and assessing accuracy of land change. *Remote Sens. Environ.* **2014**, *148*, 42–57. [[CrossRef](#)]
16. CILSS. *Landscapes of West Africa—A Window on a Changing World*; U.S. Geological Survey EROS: Garretson, SD, USA, 2016.
17. Cotillon, S.E. *West Africa Land Use and Land Cover Time Series*; U.S. Geological Survey: Sioux Falls, SD, USA, 2017.
18. Cotillon, S.E.; Mathis, M.L. *Mapping Land Cover through Time with the Rapid Land Cover Mapper—Documentation and User Manual*; U.S. Geological Survey: Sioux Falls, SD, USA, 2017.
19. Bartholome, E.; Belward, A.S. GLC2000: A new approach to global land cover mapping from Earth observation data. *Int. J. Remote Sens.* **2005**, *26*, 1959–1977. [[CrossRef](#)]
20. Arino, O.; Gross, D.; Ranera, F.; Leroy, M.; Bicheron, P.; Brockman, C.; Defourny, P.; Vancutsem, C.; Achard, F.; Durieux, L. GlobCover: ESA service for global land cover from MERIS. In Proceedings of the 2007 IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2007), Barcelona, Spain, 23–27 July 2007.
21. Arino, O. GlobCover 2009. European Space Agency. Available online: http://epic.awi.de/31046/1/Arino_et_al_GlobCover2009-a.pdf (accessed on 25 April 2018).
22. UCL-Geomatics. Land Cover CCI Product User Guide Version 2.0. Available online: http://maps.elie.ucl.ac.be/CCI/viewer/download/ESACCI-LC-Ph2-PUGv2_2.0.pdf (accessed on 10 April 2017).
23. Chen, J.; Chen, J.; Liao, A.; Cao, X.; Chen, L.; Chen, X.; He, C.; Han, G.; Peng, S.; Lu, M. Global land cover mapping at 30 m resolution: A POK-based operational approach. *ISPRS J. Photogramm. Remote Sens.* **2015**, *103*, 7–27. [[CrossRef](#)]

24. Fritz, S.; See, L.; McCallum, I.; You, L.; Bun, A.; Moltchanova, E.; Duerauer, M.; Albrecht, F.; Schill, C.; Perger, C. Mapping global cropland and field size. *Glob. Chang. Biol.* **2015**, *21*, 1980–1992. [[CrossRef](#)] [[PubMed](#)]
25. Xiong, J.; Thenkabail, P.S.; Tilton, J.C.; Gumma, M.K.; Teluguntla, P.; Oliphant, A.; Congalton, R.G.; Yadav, K.; Gorelick, N. Nominal 30-m cropland extent map of continental Africa by integrating pixel-based and object-based algorithms using sentinel-2 and Landsat-8 data on Google earth engine. *Remote Sens.* **2017**, *9*, 1065. [[CrossRef](#)]
26. Chen, Y.; Zhou, Y.N.; Ge, Y.; An, R.; Chen, Y. Enhancing Land Cover Mapping through Integration of Pixel-Based and Object-Based Classifications from Remotely Sensed Imagery. *Remote Sens.* **2018**, *10*, 77. [[CrossRef](#)]
27. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
28. See, L.; Laso Bayas, J.C.; Schepaschenko, D.; Perger, C.; Dresel, C.; Maus, V.; Salk, C.; Weichselbaum, J.; Lesiv, W.; McCallum, I.; et al. LACO-Wiki: A new online land cover validation tool demonstrated using GlobeLand30 for Kenya. *Remote Sens.* **2017**, *9*, 754. [[CrossRef](#)]
29. Townshend, J.R.G.; Huang, C.; Kalluri, S.N.V.; Defries, R.S.; Liang, S.; Yang, K. Beware of per-pixel characterization of land cover. *Int. J. Remote Sens.* **2000**, *21*, 839–843. [[CrossRef](#)]
30. Castanedo, F. A review of data fusion techniques. *Sci. World J.* **2013**, *2013*, 1–19. [[CrossRef](#)] [[PubMed](#)]



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).