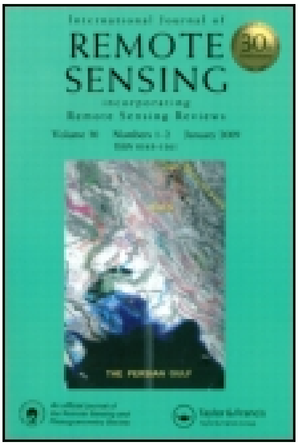


International Journal of Remote Sensing
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tres20

Comparison of global and regional land cover maps with statistical information for the agricultural domain in Africa

Steffen Fritz [a], Linda See [b] & Felix Rembold [c]

[a] International Institute for Applied Systems Analysis (IIASA), Am Schlossplatz 1, A-2361, Laxenburg, Austria

[b] School of Geography, University of Leeds, University Road, Leeds, LS2 9JT, UK

[c] Joint Research Centre of the European Commission, TP 266Via Enrico Fermi 2749, I-21027, Ispra (VA), Italy

Published online: 14 May 2010.

To cite this article: Steffen Fritz , Linda See & Felix Rembold (2010): Comparison of global and regional land cover maps with statistical information for the agricultural domain in Africa, International Journal of Remote Sensing, 31:9, 2237-2256

To link to this article: http://dx.doi.org/10.1080/01431160902946598

International Journal of Remote Sensing
Vol. 31, No. 9, 10 May 2010, 2237–2256

Taylor & Francis
Taylor & Francis Group

Comparison of global and regional land cover maps with statistical information for the agricultural domain in Africa

STEFFEN FRITZ*†, LINDA SEE‡ and FELIX REMBOLD§
†International Institute for Applied Systems Analysis (IIASA), Am Schlossplatz 1, A-2361 Laxenburg, Austria
‡School of Geography, University of Leeds, University Road, Leeds, LS2 9JT, UK
§Joint Research Centre of the European Commission, TP 266Via Enrico Fermi 2749, I-21027 Ispra (VA), Italy

(*Received 19 December 2007; in final form 22 August 2008*)

Achieving food security in particular in Africa continues to pose a major challenge to humankind. It is clear that the future agricultural potential of Africa plays a critical role in meeting this challenge. Although crop yield can be estimated with a degree of reliability using a limited sample of ground observations, the exact crop acreage and the spatial distribution are rarely available. Even though remote sensing offers the ability to produce a rapid and up-to-date land use and land cover database for agricultural monitoring, there are only a few countries in Africa where higher resolution satellite data such as Landsat have been used for land cover map production at the national level. However a number of global products have been produced which contain information on cropland extent. This paper will outline a comparison of four sources of land cover data to determine which product is the most suitable for agricultural monitoring and for the subsequent development of a crop mask. The land cover products used are the Global Land Cover Map (GLC-2000), the Moderate Resolution Imaging Spectroradiometer (MODIS) land cover product (MOD12V1), the SAGE (Center for Sustainability and the Global Environment) and the AFRICOVER dataset from the Food and Agriculture Organisation (FAO). Both the GLC-2000 and MODIS land cover products are at a resolution of 1 km^2 while AFRICOVER, based on visual interpretation of 30-m resolution Landsat images, is available at a much finer resolution. The four land cover products are first aggregated to the same resolution so that they can be compared. The legend categories of the four land cover products in this study are reconciled using the method developed in Fritz and See [Fritz, S. and See, L., 2005, Comparison of land cover maps using fuzzy agreement. *International Journal of Geographic Information Science*, **19**, pp. 787–807.] and See and Fritz [See, L.M. and Fritz, S., 2005, A user-defined fuzzy logic approach to comparing global land cover products. In *14th European Colloquium on Theoretical and Quantitative Geography*, 9–12 September 2005, Lisbon, Portugal.] that allows overlap between legend definitions to be taken into account. Once the legend definitions between the different land cover products are reconciled, the maps are then compared with national and sub-national statistics. Analysis is undertaken at both continental and national scales as well as sub-national for Sudan and Eritrea. The study generally concludes that MODIS has the tendency to underestimate cropland cover when compared with FAO

*Corresponding author. Email: fritz@iiasa.ac.at

International Journal of Remote Sensing
ISSN 0143-1161 print/ISSN 1366-5901 online
http://www.tandf.co.uk/journals
DOI: 10.1080/01431160902946598

statistics or AFRICOVER data, whereas GLC-2000 tends to overestimate cropland cover in those countries that are located at the northern transition zone of subtropical shrubland and semi-desert areas. In this area MODIS and SAGE show a relatively similar cropland distribution. Even though the SAGE database has been calibrated with national statistics, it does not perform better than the other two datasets overall, and has highlighted the fact that the SAGE data show regional weaknesses and should be replaced in certain regions by more recent datasets such as GLC-2000 and MODIS, or ideally by a hybrid product that combines the best of the three products, depending upon the region and country.

1. Introduction

Providing a sufficient supply of food and water over large areas of Africa still poses a major challenge to humankind. A report covering the Millennium Development Goals shows that agriculture plays an important role in achieving these challenges (UN Millennium Project 2005). To improve the current food supply, knowledge about crop potential is a crucial input. Although crop yield can be estimated with a certain degree of reliability using a limited sample of ground observations, the exact crop acreage and the spatial distribution are rarely available. The lack of this kind of data poses an obstacle in the development of models that provide insight into actual food potential. Even though remote sensing offers the ability to produce a rapid and up-to-date land use and land cover database for agricultural monitoring, there are only few countries in Africa where higher resolution satellite data such as Landsat are used for land cover map production at the national level, e.g. the AFRICOVER project (Food and Agriculture Organisation (FAO) 1998). Accurate information within the agricultural domain is particularly important for crop monitoring for the purpose of food security. Several institutions monitoring crop status in food-insecure countries and in an early warning systems context are highly interested in crop distribution and crop extent maps. Also the inter-annual variations of crop areas are of big interest and very difficult to produce in highly fragmented rural landscapes as commonly found in many African countries. For example, the FOOD-SEC action of the European Commission Joint Research Centre is monitoring agricultural vegetation in vulnerable countries around the world using mainly remote sensing data. One of the main techniques used consists of extracting vegetation index temporal profiles from low resolution satellite images in combination with land cover/land use maps to link the extracted indicators with the observed crops.

This paper will outline a comparison of four sources of land cover data to determine which product is the most suitable for agricultural monitoring and for the subsequent development of a crop mask, an important input to both food security and monitoring of agricultural expansion. Whereas earlier papers have considered a global comparison of land cover products (Fritz and See 2005, 2008, See and Fritz 2005), this paper focuses on the agricultural domain in Africa. The land cover products used are the Global Land Cover Map (GLC-2000) (Fritz *et al.* 2003), the Moderate Resolution Imaging Spectroradiometer (MODIS) land cover product (MOD12V1) (Friedl *et al.* 2002), SAGE (Center for Sustainability and the Global Environment) (Leff *et al.* 2004) and the AFRICOVER dataset from the FAO (1998). Both the GLC-2000 and MODIS land cover products are at resolution of 1 km^2 while AFRICOVER, based on visual interpretation of 30-m resolution Landsat images, is available at a much finer resolution.

The four land cover products are first aggregated to the same resolution so they can be compared. Map comparisons usually involve a Boolean or crisp approach, which means that legend categories from the different maps, which are often very different, are matched using a one-to-one mapping regardless of obvious incompatibilities. The resulting map comparison then shows areas with 100% agreement or disagreement. The problem with this type of approach is that the user is not certain whether the disagreement is real or just a function of semantic differences in the legend definitions. The legend categories of the four land cover products in this study are reconciled using the method developed in Fritz and See (2005, 2008) and See and Fritz (2005), which allows overlap between legend definitions to be taken into account. This is particularly important in this study as the legend definitions that refer to the agricultural domain are not entirely compatible and three products use a range (e.g. agriculture 20% to 50%) rather than a single value.

Once the legend definitions between the different land cover products are reconciled, the maps are then compared with national and sub-national statistics. Analysis is undertaken at both continental and national scales as well as sub-national for Sudan and Eritrea. Sudan is the AFRICOVER country with the largest area of arable land while for Eritrea sub-national statistics have been made available reporting detailed agricultural statistics for smaller administrative units. The provincial statistical analysis also serves to cross-check the AFRICOVER dataset with official sub-national statistics from the Sudanese and Eritrean Ministry of Agriculture in order to have more confidence in the quality of this product. The AFRICOVER dataset is then used as a reference dataset against which the other three products are compared using a fuzzy logic approach. The percentage disagreement between each of the three global land cover products, i.e. GLC-2000 for Africa, MODIS and SAGE, is then calculated as outlined in previous studies (e.g. Fritz and See 2005). The methodology is then further extended to calculate the disagreement in terms of omission and commission using AFRICOVER as a reference dataset. The results are discussed and recommendations are made regarding the suitability of the different land cover products for creating an agricultural mask for the AFRICOVER countries as well as those outside of the AFRICOVER area.

2. The land cover products and FAO agricultural statistics

The Global Land Cover 2000 (GLC-2000) is a global product for the baseline year 2000, a reference year for environmental assessment (e.g. Millennium Ecosystems Assessment). This dataset was created in collaboration with partners around the world (Fritz *et al.* 2003, Bartholomé and Belward 2005) and is based on the VEGETATION sensor onboard the Satellite Pour l'Observation de la Terre (SPOT)-4 satellite. GLC-2000 was developed using a bottom-up approach in which more than 30 research teams contributed to 19 regional windows, where the regional legends used the Land Cover Classification System (LCCS) as a common language to produce 22 global classes (Di Gregorio and Jansen 2000, Fritz *et al.* 2003). The GLC-2000 for Africa was produced on a regional level with 27 different classes which are more refined and detailed, but they map directly onto the 22 global classes. Therefore the regional African part of Global Land Cover 2000 is used in this study and is hereafter referred to simply as GLC-2000.

The MODIS land cover product from Boston University (MOD12Q1 V004, 1 km) was created using the Moderate Resolution Imaging Spectoradiometer instrument on the NASA Terra Platform using data from the period mid-October 2000 to mid-October 2001. The MODIS land cover dataset uses all 17 classes of the IGBP (International Global Biosphere Project) legend (Loveland *et al.* 2000), and unlike GLC-2000, a global classification approach has been used. Both the GLC-2000 and the MODIS datasets are currently the most recent global land cover products available.

A global cropland dataset has been produced by the Center for Sustainability and the Global Environment (SAGE) at the University of Wisconsin. It is considered to be the global standard dataset for agricultural applications since it was calibrated with the use of national statistics (Ramankutty and Foley 1998) and has been refined to serve as a basis for crop specific agricultural areas globally (Leff *et al.* 2004). SAGE uses a data fusion technique to integrate remotely sensed data derived from the Global Land Cover characteristics database (Loveland *et al.* 2000) and administrative-unit level inventory data (Ramankutty and Foley 1998). However, it is based on relatively old Advanced Very High Resolution Radiometer (AVHRR) data (Loveland *et al.* 2000), which has a number of reported problems such as poor sensor calibration, variable angular-induced pixel sizes (Miura *et al.* 2000) and in particular, high geo-registration errors (Bartholomé and Belward 2005). The accuracy of this dataset, especially in Africa, is therefore questionable. The methodology proposed in this paper is used to determine the validity of this dataset in the AFRICOVER countries by focusing on the spatial disagreement with the AFRICOVER dataset.

AFRICOVER began as an FAO project funded by the Italian government to help countries in Africa set up a geo-referenced database on land cover using visual on-screen interpretation of Landsat data in combination with ground data and expert knowledge. It currently provides detailed, baseline agricultural land use information for 10 countries in Africa: Burundi, Democratic Republic of Congo, Egypt, Eritrea, Kenya, Rwanda, Somalia, Sudan, Tanzania and Uganda (FAO 1998). Most of the Landsat images used for these countries were acquired in the period between 1995 and 1999 and have been interpreted at a scale of 1:200 000 or 1:100 000 respectively for large or small countries. A new land cover classification system (LCCS) was developed by the project and has become a standard for Land Cover products (Di Gregorio and Jansen 2000).

Agricultural statistics for a single country can be extracted by examining each single polygon belonging to a given LCCS class. For single code polygons, the area is 100%. Polygons with more than one LCCS code contain mixed classes. In the case of two LCCS codes, the first class is assigned 60% while the second is given 40% of the polygon area. Similarly, for polygons with three codes, the percentages 40, 30 and 30 are used respectively. In addition there are two special cases: (i) scattered clustered agriculture, which is assigned 35% for 2nd and 3rd codes; and (ii) scattered isolated agriculture, which is assigned 15% for codes 2 and 3. Some codes are then adjusted to make sure the total assigned is 100%. AFRICOVER has recently been compared to GLC-2000 by Torbick *et al.* (2005), who found a 54% agreement overall for agriculture. The vision is to extend the AFRICOVER project to the whole of Africa upon further funding and donor support. The same methodology has recently been applied to Libya under a parallel FAO initiative, and is now underway in Senegal.

FAOSTAT (2005) is an online, multilingual database that holds statistics for many different countries, including agricultural time series data. Individual countries provide their national statistics each year for this official reference database. The FAO database defines arable land as land that is under temporary crops (double-cropped areas are counted only once), temporary meadows for mowing or pasture, land under market and kitchen gardens, and land temporarily fallow (less than five years). Any land that is abandoned as a result of shifting cultivation is not included in this category. Note that the definition of arable land does not indicate areas that have potential for cultivation. Since permanent crops are excluded in some of the definitions of the land cover products used in this study, only countries where permanent crops/tree crops play a negligible role are considered. Sub-national crop statistics from the Sudanese and Eritrean Ministries of Agriculture were obtained, which gives information on crop statistics at a provincial or State level (Maragan 2001).

3. Methodology

The methodology consists of first aggregating the different land cover maps to a compatible resolution. The four sets of legend classes are then reconciled so that comparison of the overall agricultural areas for the different datasets together with official national and sub-national statistics can be undertaken. Finally, fuzzy logic is used to facilitate a spatial comparison between the GLC-2000, MODIS and SAGE datasets in relation to the AFRICOVER database for African countries that record a negligible proportion of tree crops, as explained previously. Hotspots or clusters of disagreement are identified between each land cover product and AFRICOVER. An overview of the datasets and the processes involved is shown in figure 1.

3.1 *Land cover aggregation*

To facilitate comparison, the different land cover products were aggregated to the resolution of the SAGE database, i.e. 5 minutes or approximately 10 km. This aggregation will also compensate for some geo-location problems (Klein *et al.* 1993). To aggregate AFRICOVER to 5 minutes, cropland in the 10 000 individual pixels was added up and divided by 10 000 to arrive at an average percentage cropland. This aggregation was undertaken for both the minimum and maximum value of each class range. In this way a minimum and maximum cropland fraction image is produced that records the percentage area of cropland on a scale from 0–100% found in each grid of 100 km^2 and thereby incorporates the uncertainty present in the definitions. The AFRICOVER fraction GRID is shown in figure 2.

The aggregation operation was then applied to both MODIS and GLC-2000 in the same way that the AFRICOVER dataset was aggregated. The following classes with their definitions were aggregated: cultivated and managed (from the GLC-

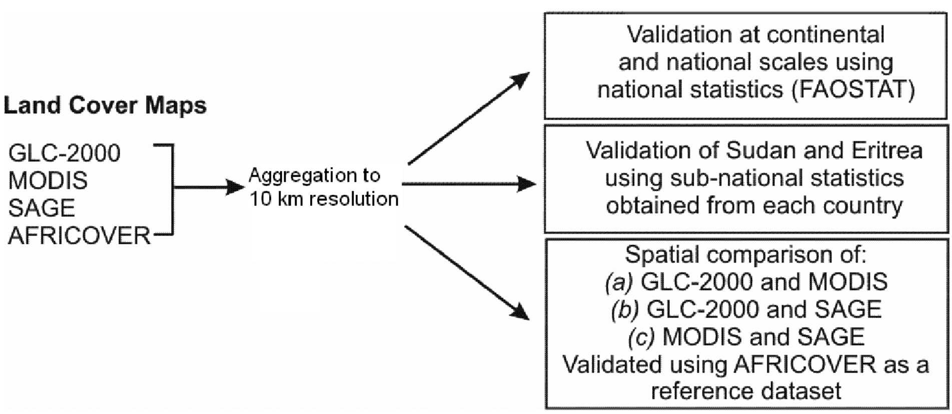


Figure 1. The datasets used in the study for comparison and validation.

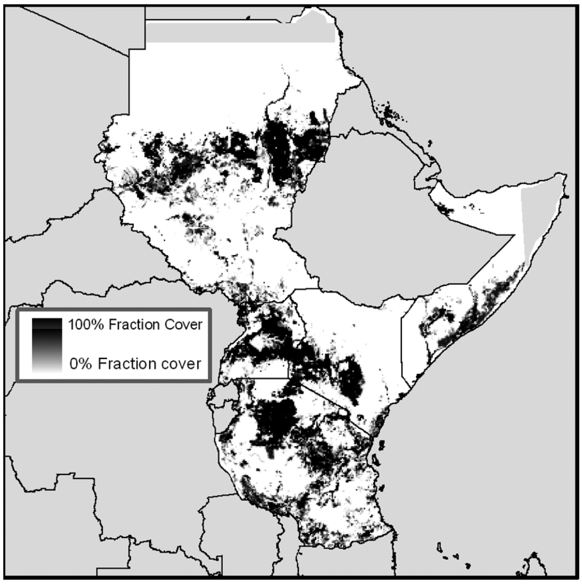


Figure 2. Example of AFRICOVER fraction image (AFRICOVER Max).

2000), cropland/forest mosaic (from GLC-2000), cropland/natural mosaic (from GLC-2000), irrigated agriculture (from GLC-2000), cropland (from MODIS) and cropland/natural vegetation mosaic (from MODIS). Similarly, both a minimum and maximum cropland cover map were created for MODIS and GLC-2000 using the definitions of the different classes (e.g. for the GLC-2000 class 'cultivated and managed', a minimum of 50% and a maximum of 100% was assigned).

3.2 *Reconciliation of legend classes*

Table 1 contains the different definitions for agricultural land from the four different land cover products. In general, researchers tend to use only one percentage value of cropland when making this type of comparison (e.g. see Hannerz and Lotsch 2008). In contrast, the methodology in this study considers the whole range of cropland values. Disagreement is, therefore, only recorded outside this range. Based on the definitions in table 1, a legend look-up table that links the four land cover products can be derived. For example, table 2 contains the look-up table for GLC-2000 and AFRICOVER. Where there is any degree of overlap, complete agreement is assumed for simplicity and marked with an X. Similar look-up tables for agreement are derived between each pair of land cover products.

The presence of tree crops in the SAGE database means that the approach taken in Fritz and See (2005) would result in the entire SAGE cropland database being mapped onto all the shrub and tree cover of other land cover maps. As a result there would hardly be any disagreement recorded due to the large degree of overlap. Therefore, an alternative is to use only those countries that have a negligible amount of tree crops, i.e. less than 5% of arable area, so that a more reasonable comparison is possible.

Table 1. Definitions of cropland including the percentage cover in the relevant legend types from the different land cover products.

Legend type	Definition	Cropland (%)
MODIS	Croplands are lands covered with temporary crops followed by harvest and a bare soil period (e.g. single and multiple cropping systems). Note that perennial woody crops will be classified as the appropriate forest or shrub land cover type.	15–60
	Cropland/natural vegetation mosaics are lands with a mosaic of croplands, forests, shrublands and grasslands in which no one component comprises more than 60% of the landscape.	60–100
GLC-2000 for Africa	Croplands: Areas with over 50% cultures or pastures. Regions of intensive cultivation and/or sown pasture fall in this class.	50–100
	Mosaic Forest-Cropland: The vegetation found here is formed by a complex of secondary regrowth, fallow, home gardens, food crops and village plantations.	15–50
	Mosaic Cropland – Natural Vegetation: At the south of the Sahelian belt, the croplands are mixed with natural vegetation and represent up to 30% of the cover.	15–30
	Irrigated agriculture: Agriculture depending on artificial water supply.	100
SAGE	Arable land (including harvested cropland, crop failure, temporarily fallow or idle land and cropland used temporarily for pasture) and land under permanent crops (such as cocoa, coffee, rubber; including all tree crops except those grown for wood or timber). The harvested produce may be used for both human consumption and/or feed	Calibrated in Africa with national statistics so no range
AFRICOVER	Pure continuous fields, herbaceous crops	80–100
	Mixed continuous fields, herbaceous crops	50–80
	Scattered clustered, herbaceous crops	20–50
	Scattered isolated, herbaceous crops	10–20
	Pure continuous fields, shrub or tree crops	80–100
	Mixed continuous fields, shrub or tree crops	50–80
	Scattered clustered, arable, shrub or tree crops	20–50
	Scattered isolated, arable, shrub or tree crops	10–20

3.3 *Comparison of agricultural statistics (from FAO and sub-national censuses) with land cover products*

To compare the official agricultural statistics from FAO with the cropland areas in the land cover datasets, the areas of cropland by country were extracted. For both AFRICOVER and SAGE, a single value was obtained per country. In contrast, both GLC-2000 and MODIS provide a minimum and maximum value. To identify the dataset that coincides most closely with these official statistics, the root mean squared error (RMSE) was calculated as follows:

Table 2. Example of matrix between AFRICOVER and GLC-2000.

	AFRICOVER Category			
	Pure continuous fields, arable (80–100% cover)	Mixed continuous fields, arable (50–80% cover)	Scattered clustered, arable (20–50% cover)	Scattered isolated, arable (10–20% cover)
GLC-2000 category				
Croplands (50–100% cover)	X	X		
Mosaic forest-cropland (15–50% cover)			X	X
Mosaic cropland/ natural vegetation (15–30% cover)			X	X
Irrigated agriculture (100% cover)	X			

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (O_i - P_i)^2}{n}} \tag{1}$$

where O_i is the crop area as reported by FAO, P_i is the value derived from the land cover maps and n is the number of countries. The RMSE was also calculated for Sudan and Eritrea at sub-national level, where n would then be the number of sub-national units, e.g. states or provinces. The RMSE is a standard measure for assessing accuracy in spatial analysis (Siska and Hung 2001), which penalizes larger errors. This measure is therefore useful when national and sub-national statistics from different datasets are compared with official values.

3.4 *Creation of percentage disagreement maps*

The final step is the spatial comparison of the land cover products; in particular the creation of percentage disagreement maps between AFRICOVER and the other land cover products. Each pair of land cover maps was compared on a pixel by pixel basis using the look-up tables described in §3.2. An X in the look-up table (see table 2) signifies an overlap in the definitions of the legend classes of a pair of given maps so the percentage disagreement is considered to be 0%. For all other pixel comparisons, the percentage difference is mapped onto a fuzzy set like that shown in figure 3, to denote degrees of difference, which are then mapped spatially. Hotspot maps of disagreement between MODIS and SAGE, MODIS and GLC-2000, and GLC-2000 and SAGE were then produced.

A more refined regional analysis was then undertaken for eastern African countries (Sudan, Uganda, Somalia and Tanzania) for which AFRICOVER is available. AFRICOVER is used as a reference against which the others are compared since it is based on a visual interpretation of high-resolution images and ground information. Moreover, it was then possible to calculate the areas of omission and commission

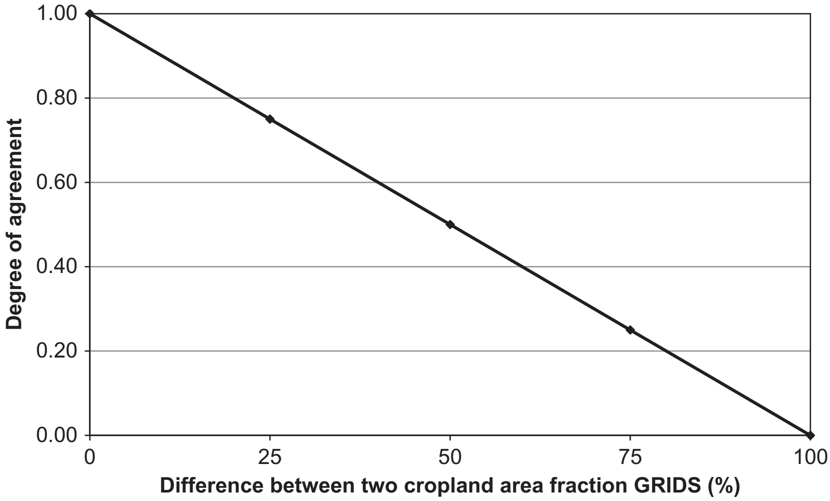


Figure 3. Fuzzy set showing degree of agreement for the percentage difference between cropland areas.

(Congalton 1991). The area of omission is calculated as the overall area of disagreement in areas where AFRICOVER records a higher cropland fraction. This also includes the case where AFRICOVER records a certain cropland fraction and the other dataset does not have any crops present. Conversely, the area of commission is calculated where AFRICOVER records a smaller cropland fraction or zero cropland fraction compared to the other dataset. In this evaluation the minimum and maximum cropland fraction is taken into account. For example, if a given pixel in AFRICOVER records a minimum and maximum cropland fraction of 20% and 40% respectively, while the other dataset (e.g. GLC-2000) records 50% and 100% for the minimum and maximum values, then the percentage disagreement is 10%, which corresponds to a commission area of 10 km^2 (i.e. 10% of 100 km^2).

4. Results

4.1 *Continental, national and sub-national comparison*

The total crop areas for Africa derived by summing the FAO national statistics along with estimates from the different land cover products and the SAGE database are given in figure 4. Unsurprisingly the SAGE database gives total values that are close to the totals from the FAO statistics because this dataset has been calibrated at a continental level using the best correlation (R^2). Any differences are due to the fact that SAGE was calibrated using older national statistics, i.e. 1990. The GLC-2000 crop areas summed using the minimum value of the range also produce values at continental level that are similar to the overall continental total derived from FAO. The minimum and maximum crop areas from MODIS both underestimate total crop area at a continental level.

4.1.2 Country statistics for selected African countries. Figure 5 provides the same comparison but at national level for selected countries in Africa, representing a range of different ecosystems across the continent. It is clear from looking at the graph that

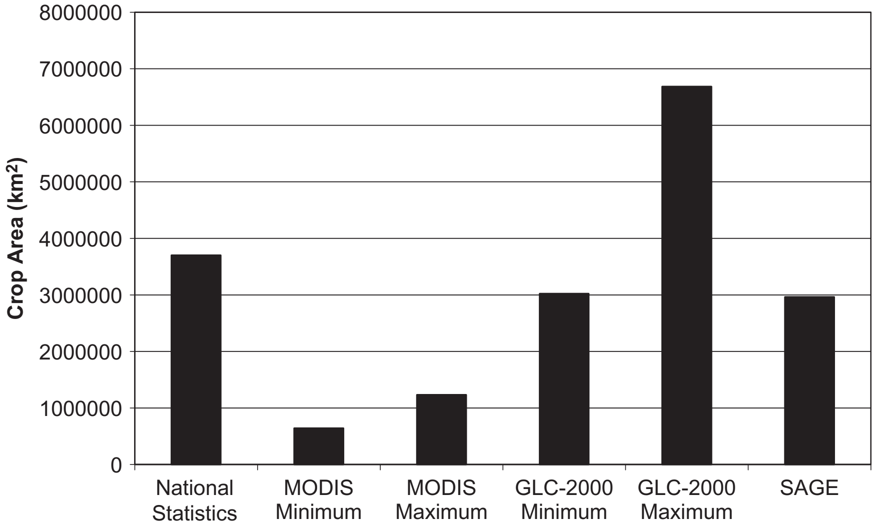


Figure 4. Total crop area from the national statistics (year 2000) compared to the land cover products for Africa.

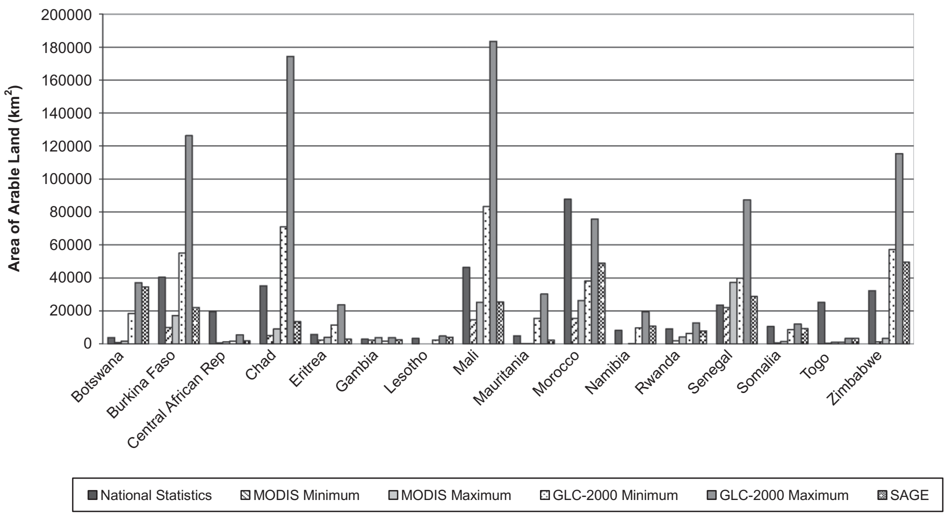


Figure 5. Area of arable land per country (reference year 2000).

there is a large amount of difference between the estimates; this highlights the high amount of uncertainty when comparing with national statistics. Table 3 provides the RMSE for each of the countries and shows that the best fit is found using the GLC-2000 minimum figures followed by SAGE. The SAGE result is not surprising because this dataset was calibrated using national statistics. The MODIS maximum and minimum both result in RMSE values that are slightly higher than the GLC-2000 minimum and SAGE; however, they are relatively close together. The GLC-2000

Table 3. RMSE of cropland area comparing each land cover product to national FAO statistics.

Land cover type	RMSE (km^2)
GLC-2000 minimum	21064
GLC-2000 maximum	76802
SAGE	25109
MODIS minimum	27787
MODIS maximum	36504

minimum and maximum, on the other hand, are further apart, with the GLC-2000 maximum map producing the highest RMSE.

4.1.3 Sub-national comparison for Sudan. For each administrative unit within Sudan we selected the year in which the majority of the Landsat TM data scenes for the AFRICOVER classification had been used (mainly 1996 and 1997). Figure 6 provides a comparison of cropland area estimates from the different land cover products and the official statistics (from the Sudanese Ministry of Agriculture) for the different states (sub-national units) of Sudan, while table 4 contains the overall RMSE. AFRICOVER has the lowest RMSE and matches the sub-national statistics most closely. Both SAGE and MODIS have higher RMSEs of roughly the same magnitude while GLC-2000 has the highest errors.

4.1.4 Sub-national comparison for Eritrea. For Eritrea, nearly all Landsat scenes were interpreted for the AFRICOVER classification from the year 1999. We therefore chose sub-national statistics from the same time period. Figure 7 provides a

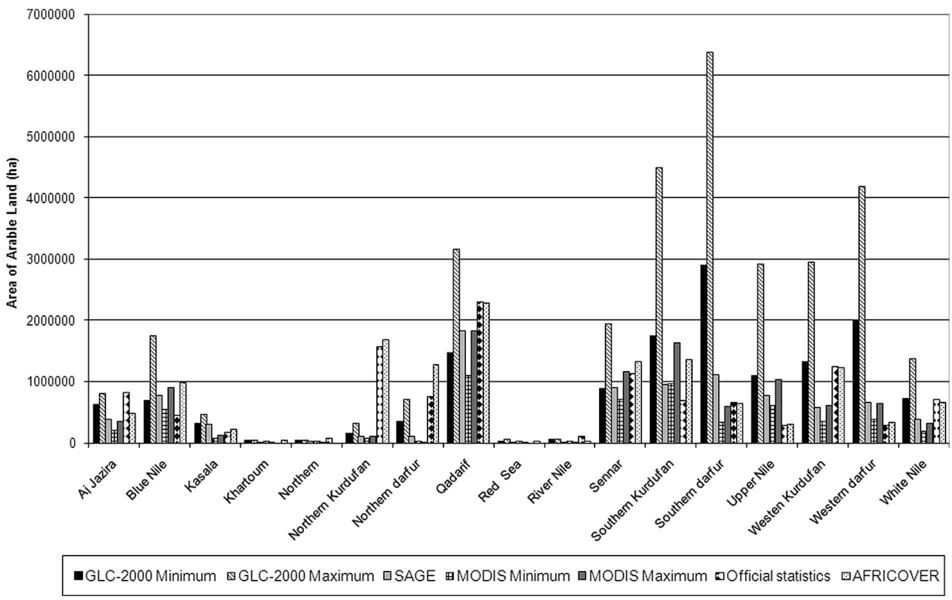


Figure 6. Area (in ha) of arable land by state in Sudan.

Table 4. RMSE for cropland area comparing each land cover product to the official sub-national statistics (Ministry of Agriculture) for Sudan and Eritrea.

Land cover type	Sudan RMSE (km^2)	Eritrea RMSE (km^2)
GLC-2000 minimum	8672	1525
GLC-2000 maximum	21335	4155
SAGE	5038	1163
MODIS minimum	6034	660
MODIS maximum	5679	770
AFRICOVER	2669	641

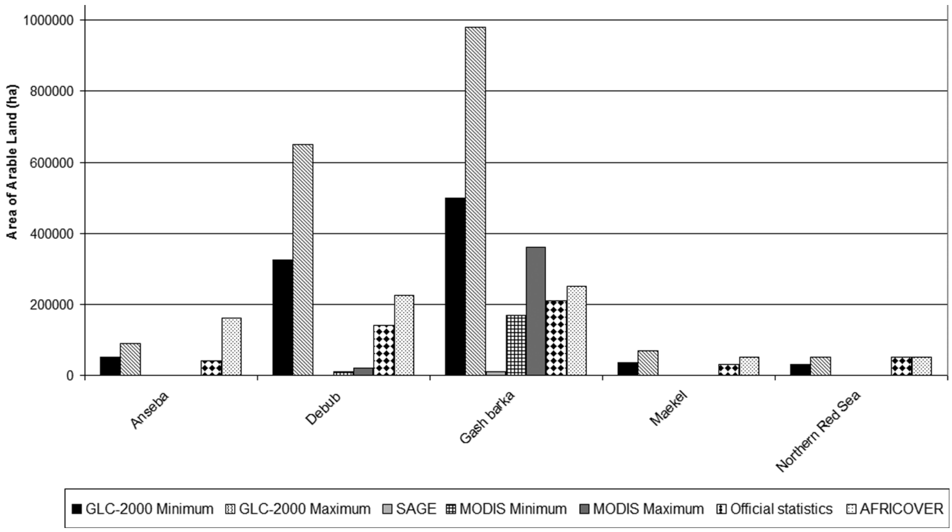


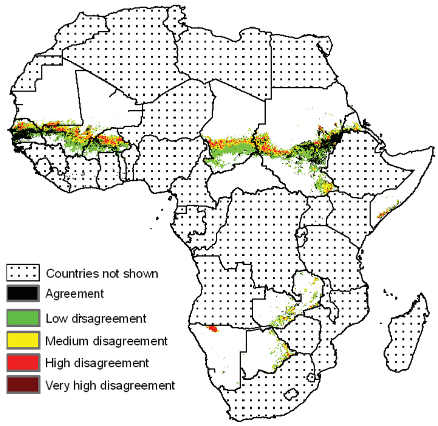
Figure 7. Area of arable land (in ha) by Zobas (sub-national level) in Eritrea.

comparison of cropland area estimates from the different land cover products and the sub-national agricultural statistics at the Zoba level (provided by the Ministry of Agriculture) for all Zobas (sub-national units) except for South Red Sea, where agricultural areas are very marginal. In terms of RMSE we observe similar patterns as for Sudan, with AFRICOVER performing the best followed by MODIS and SAGE for this country.

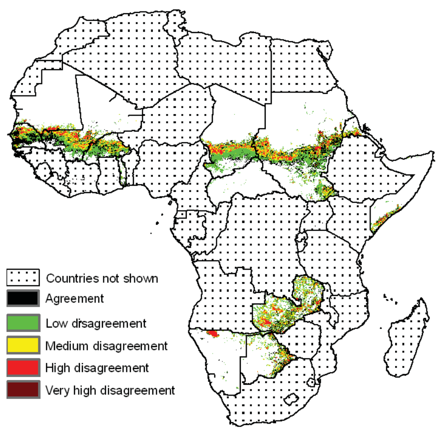
4.2 *Spatial disagreement for selected African countries*

The previous section dealt with a comparison of how the different land cover products match official statistics, but this does not tell us anything about the accuracy of the spatial distribution of cropland. Figure 8(*a*) shows the spatial disagreement between GLC-2000 and MODIS land cover products. High areas of disagreement can be seen in Mali, north-eastern Burkina Faso, western Chad and western Sudan. The GLC-2000 and the SAGE datasets are compared in figure 8(*b*). Similar patterns to those shown in figure 8(*a*) can be found but with less severe disagreement in Burkina Faso

(*a*)



(*b*)



(*c*)

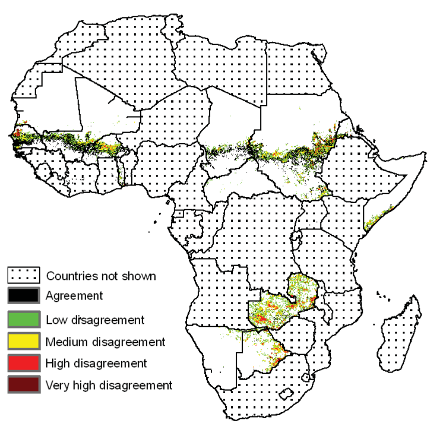


Figure 8. Agreement and disagreement in the agricultural domain between (*a*) GLC-2000 and MODIS; (*b*) GLC-2000 and SAGE; and (*c*) MODIS and SAGE land cover maps. White areas are where both datasets denote no cropland.

and more in Zambia. Finally, figure 8(*c*) shows the spatial disagreement when comparing MODIS and SAGE. There are few areas of severe disagreement but in Zambia, there are similar patterns to those found in the GLC-2000 and SAGE comparison.

When comparing the patterns shown on the spatial disagreement maps with national statistics, there is an indication that GLC-2000 in countries such as Burkina Faso, Chad, Senegal, Zimbabwe and Sudan (see figure 4) records cropland in areas where there is either very little or no cropland since even the minimum cropland value for those countries is by far higher than the official national statistics. For the sub-national statistics in Sudan, the same phenomena can be observed, particularly in the states of Southern Kordufan, Southern Darfur, the Upper Nile and Western Darfur. On the other hand, in Zambia, both SAGE and GLC-2000 are closer to the national statistics and MODIS does not show large areas of cropland, possibly in those areas where both the GLC-2000 and SAGE record cropland. Final conclusions about these countries, however, are difficult to make because the official national FAO statistics are also prone to error and must therefore be treated with care. The patterns indicated are simply general trends and the true spatial distribution of cropland in these areas should be examined in more detail. For Sudan and Eritrea the conclusions are already more robust as sub-national statistics were considered and AFRICOVER data were available for verification. We therefore present results of a more detailed analysis in the next two sections for those countries that are covered by AFRICOVER and have a high proportion of cropland, namely Sudan, Tanzania, Uganda, Kenya and Somalia.

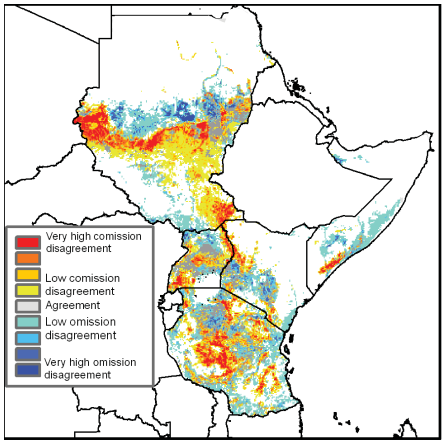
4.3 *Spatial disagreement of omission and commission errors*

Figures 9(*a*)–9(*c*) show a direct spatial comparison for the AFRICOVER countries. Figure 9(*a*) shows the area of spatial disagreement of omission and commission for AFRICOVER and GLC-2000, figure 9(*b*) is the comparison for AFRICOVER and the SAGE dataset, and figure 9(*c*) is the comparison with MODIS. In Sudan all three land cover products miss a high proportion of cropland in the northern part of the country (shown in blue tones) while GLC-2000 shows large areas of commission in the central western portion. A high proportion of these areas of commission lie in Western and Southern Darfur where the provincial statistics also record three times the cropland area indicated by the GLC-2000 minimum. For Uganda the GLC-2000 shows little omission and commission error occurring in the central and south-western part (red tones), whereas both MODIS, and in particular SAGE, record larger areas of omission in the central northern part. SAGE shows high areas of commission in the south-west around Lake Victoria. All datasets show high areas of omission in Tanzania, whereas MODIS shows hardly any in terms of areas of commission, except a very small proportion in the east. In Kenya a similar pattern can be observed with hardly any commission errors for MODIS. In terms of omission, however, all three datasets miss cropland areas in the same places. The same pattern is also observed in Somalia.

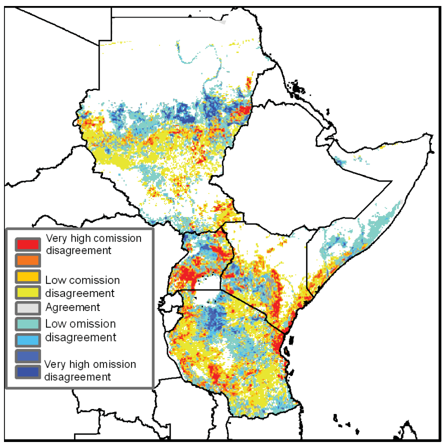
4.4 *Total area of omission and commission by country*

The total area of omission and commission for MODIS, GLC-2000 and SAGE using AFRICOVER as a reference dataset is provided in figures 10(*a*) and 10(*b*).

(*a*)



(*b*)



(*c*)

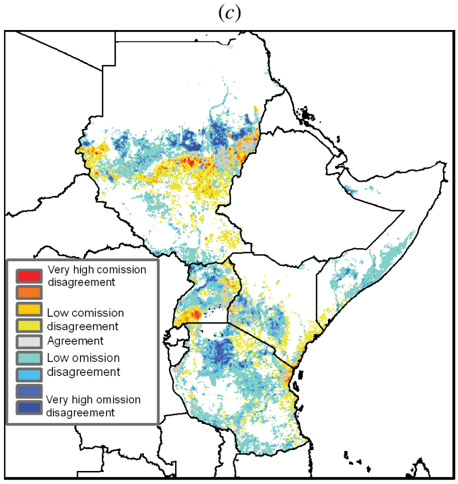


Figure 9. Areas of omission and commission comparing (*a*) AFRICOVER with the GLC-2000; (*b*) AFRICOVER with the SAGE dataset; (*c*) AFRICOVER with MODIS in the agricultural domain.

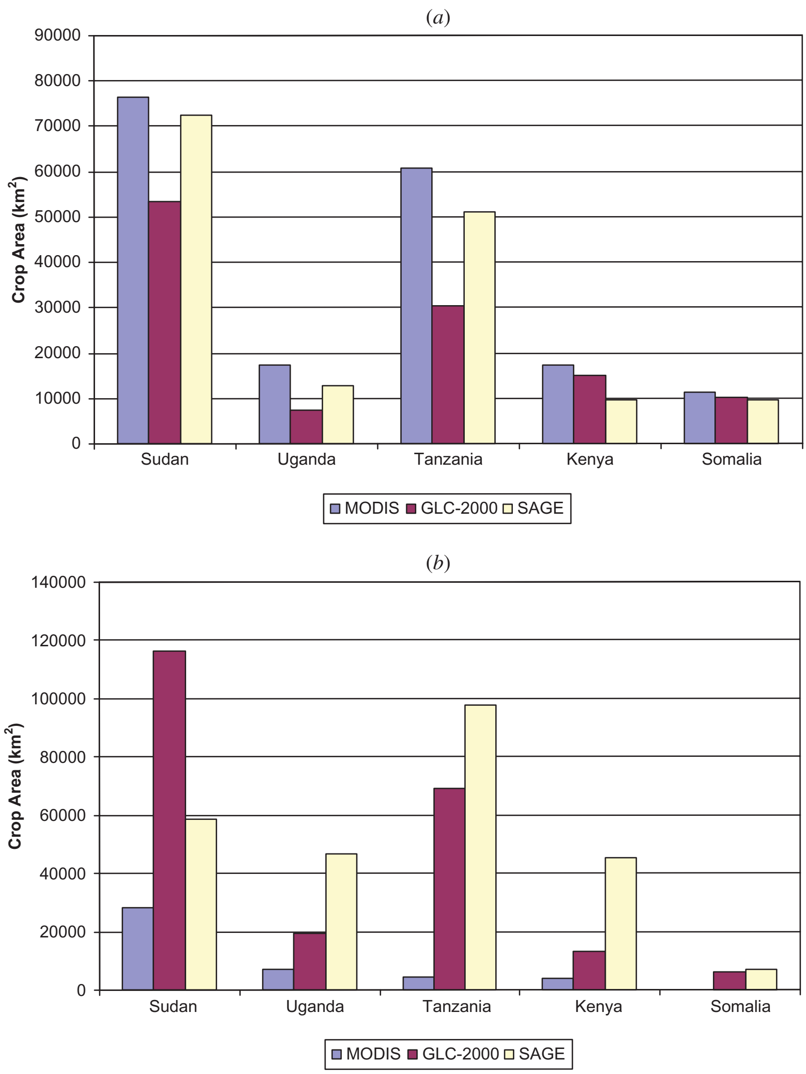


Figure 10. Areas of (*a*) omission and (*b*) commission between AFRICOVER and different land cover products.

Figure 10(*a*) shows that the GLC-2000 has the lowest omission errors for Sudan, Uganda and Tanzania, while for Kenya and Somalia the errors are greater between the different land cover products. In terms of commission, figure 10(*b*) shows that MODIS has the lowest commission errors for all five countries, yet it consistently has the highest omission errors.

5. Discussion and conclusions

Food security for early warning purposes and monitoring of agricultural expansion both require good baseline information on agriculture. This paper presented a methodology for the comparison of land cover products, in particular four recent remote sensing derived land cover maps, to determine their suitability in producing a crop mask. The overall crop area was validated with national statistics and a further sub-national analysis was undertaken for Sudan and Eritrea to understand which dataset is closest to the official statistics. The ability to undertake sub-national analyses is dependent upon the availability of crop acreage statistics. However, the analysis could be extended to other countries in the future using data from the FAO AGRO-maps project (see http://www.ifpri.org/data/gs_agromaps.htm) or by collecting more data directly within the country. AFRICOVER was used as a reference dataset because of its good performance when compared with sub-national statistics and because it was produced at a high resolution with visual interpretation and detailed ground information.

In order to have information on the disagreement of the datasets where AFRICOVER was not available, a spatial comparison of the continental datasets MODIS versus SAGE, GLC-2000 versus MODIS and GLC-2000 versus SAGE was carried out. Subsequently, a spatial comparison between the AFRICOVER database and the lower resolution land cover maps, namely MODIS, GLC-2000 and the SAGE database was undertaken. Agreement maps which use a methodology for the percentage aggregation of the land cover products was used to make an overall disagreement analysis at the resolution of the SAGE database. By focusing only on those countries with a negligible proportion of tree crops and by using tree crops in AFRICOVER with the SAGE dataset, we were able to compare the different datasets without running into problems related to incompatibility of the legends.

It has been shown that uncertainties in the cropland distribution in African countries are very high. High spatial disagreement between AFRICOVER and the other three datasets indicate that they have a number of limitations for certain applications within Africa. These preliminary results must be examined carefully as there is a certain degree of inter-annual variation of cropland area because the datasets have been produced at different times. Even though the comparison of products which followed different methodologies, were produced at different times, used different spatial scales and different definitions of cropland area is not straightforward, a number of patterns emerge when the described methodology is applied. We can generally conclude that MODIS has a tendency to underestimate cropland cover when compared with FAO statistics or AFRICOVER data, whereas GLC-2000 tends to overestimate cropland cover in those countries that are located at the northern transition zone of subtropical shrubland and semi-desert areas. In this area MODIS and SAGE show a relatively similar cropland distribution. Even though the SAGE database has been calibrated with national statistics, it does not perform better than the other two datasets overall, and has highlighted the fact that the SAGE data shows regional weaknesses and should be replaced in certain regions by more recent datasets such as GLC-2000 and MODIS or ideally by a hybrid product that combines the best of the three products, depending upon region and country. Moreover, this dataset is based on the older AVHRR sensor with a poorer sensor calibration and a certain geolocation error, and is therefore potentially inferior.

This work has also demonstrated that even though overall cropland areas in administrative units (e.g. FAO national statistics) are not so far apart, the spatial distribution of these can vary and a high uncertainty exists when a comparison is undertaken on a GRID level. This work has highlighted the need for additional development of an uncertainty layer. This will allow those areas where there is a high range (e.g. 50–100% cropland cover) to be quantified and disagreement to be considered in combination with this uncertainty layer. Further work will focus on the selection of auxiliary information to help decide which map is better for those areas where AFRICOVER is not available as well as the development of a hybrid map based on AFRICOVER, MODIS and GLC-2000, which will allow the production of a dataset at a 1-km resolution, especially with current initiatives to develop more precise national maps and the further enlargement of the AFRICOVER project.

The most conservative approach was followed in this study, allowing ranges of the legend definitions to be considered in the methodology and to focus on those countries where a direct comparison due to a low percentage of tree crops is possible. It should be pointed out that by using this type of methodology here, and considering a range rather than a specific cropland area value, a higher uncertainty arises. Furthermore, even if no disagreement was recorded using this methodology, there could still be differences between the datasets and what appears on the ground. However, these differences cannot be recorded due to the differences in legend definitions. Nevertheless, even when following this conservative approach, we still find large areas of disagreement. The official national and sub-national statistics are clearly not error-free, but do support some of the spatial patterns observed. Also, part of the differences between datasets can be explained by the acquisition dates of the imagery (late 1990s for AFRICOVER, 2000 for GLC-2000, 2000/2001 for MODIS, and late 1990s for SAGE). The difference in the acquisition date was taken into account when the land cover datasets were compared to national or sub-national statistics by choosing the corresponding years. However, when the spatial datasets are compared, there is a certain bias since inter-annual variation of cropland area could not be considered. Nevertheless these variations are minor (e.g. for most African countries less than 10% in the period between 1995 and 2000) if compared with the large differences between the three different global datasets. It is clear that both this study, and crop monitoring activities for food security analysis in general, would benefit from a better knowledge of the inter-annual change dynamics in croplands as well as the socioeconomic drivers of those changes, but it should not be forgotten that the datasets examined here are very often the only data available and that quantitative inter-annual crop acreage changes are very difficult to estimate for large areas even for countries with a much higher level of data availability. This analysis allows us to focus on problematic zones and to identify those datasets which need to be examined in more detail in the countries where large discrepancies with national statistics, together with large areas of spatial disagreement with other datasets, were identified.

Furthermore, there is the issue of purpose. In the situation where these maps are used for land cover change projection, then both commission and omission have more or less the same weight, whereas for the purpose of crop monitoring it can be advantageous to give a higher importance to omission errors. For this latter purpose the GLC-2000 may be more suitable. Even though it has higher commission errors than the other land cover products, it could be more suitable since it has generally lower omission errors.

The AFRICOVER program (as part of the Global Land Cover Network (GLCN)) is currently developing a product called 'Agricultural Trend' using the USGS Landsat archive covering the period from the late 1970s to 2000. This product will show the agricultural situation in the late 1970s and 1980s compared with the situation shown in the later AFRICOVER land cover products. Changes in agricultural area and intensity of agriculture will be shown. An interesting future study would be to compare historical area statistics for each of the countries when this product becomes available. Since AFRICOVER has detailed information on herbaceous as well as tree crops it will be possible to undertake the comparison selecting the appropriate classes depending on the defined use for cropland in the global land cover products which either includes or excludes tree crops.

Acknowledgements

The authors would like to thank the FAO who have made their AFRICOVER full resolution dataset available. Furthermore, we would like to thank the Sudanese Ministry for Agriculture as well as the Ministry of Agriculture of Eritrea for making the sub-national data available. Finally, we would like to thank Craig von Hagen for his advice and for providing the sub-national statistics of the AFRICOVER dataset.

References

Bartholomé, E. and Belward, A.S., 2005, GLC2000: a new approach to global land cover mapping from Earth Observation data. *International Journal of Remote Sensing*, **26**, pp. 1959–1977.

Congalton, R.G., 1991, A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, **37**, pp. 35–46.

Di Gregorio, A. and Jansen, L., 2000, *Land Cover Classification System: Classification Concepts and User Manual* (Rome: Food and Agriculture Organization of the United Nations).

FAO, 1998, Land cover and land use: The FAO AFRICOVER Programme. Available online at: http://www.fao.org/waicent/faoinfo/sustdev/EIdirect/EIre0053.htm, (accessed January 2008).

FAOSTAT, 2005, Agricultural data, Food and Agriculture Organization of the United Nations. Available online at: http://faostat.fao.org/faostat/collections?version=ext&hasbulk=0&subset=agriculture (accessed January 2008).

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., Baccini, A., Gao, F. and Schaaf, C., 2002, Global land cover mapping from MODIS: algorithms and early results. *Remote Sensing of Environment*, **83**, pp. 287–302.

Fritz, S. and See, L., 2005, Comparison of land cover maps using fuzzy agreement. *International Journal of Geographic Information Science*, **19**, pp. 787–807.

Fritz, S., and See, L., 2008, Identifying and quantifying uncertainty and spatial disagreement in the comparison of global land cover for different applications. *Global Change Biology*, **14**, pp. 1057–1075.

Fritz, S., Bartholomé, E., Belward, A., Hartley, A., Stibig, H.J., Eva, H., Mayaux, P., Bartalev, S., Latifovic, R., Kolmert, S., Roy, P., Agrawal, S., Bingfang, W., Wenting, X., Ledwith, M., Pekel, F.J., Giri, C., Mücher, S., de Badts, E., Tateishi, R., Champeaux, J-L. and Defourny, P., 2003, *Harmonisation, Mosaicing and Production of the Global Land Cover 2000 Database (Beta Version)* (Luxembourg: Office for Official Publications of the European Communities), EUR 20849 EN.

Hannerz, F. and Lotsch, A., 2008, Assessment of remotely sensed and statistical inventories of African agricultural fields. *International Journal of Remote Sensing*, **29**, 3787–3804.

Klein, C., Dees, M. and Peltz, D.R., 1993, *Sampling Aspects in the TREES Project: Global Inventory of Tropical Forests* (Freiburg, Germany: University of Freiburg).

Leff, B., Ramankutty, N. and Foley, J.A., 2004, Geographic distribution of major crops across the world. *Global Biogeochemical Cycles*, **18**, pp. 1–27.

Loveland, T.R., Reed, B.C., Brown, J.F., Ohlen, D.O., Zhu, Z., Yang, L. and Merchant, J., 2000, Development of a global land cover characteristics database and IGBP DISCover from 1-km AVHRR Data. *International Journal of Remote Sensing*, **21**, pp. 1303–1330.

Maragan, A.M., 2001, Time series of the main food & oil crops data (70/71–000/01). Ministry of Agriculture and Forestry, General Administration of Planning and Agricultural Economics, Vol. 5.

Miura, T., Huete, A.R. and Yoshioka, H., 2000, Evaluation of sensor calibration uncertainties on vegetation indices for MODIS. *IEEE Transactions on Geoscience and Remote Sensing*, **38**, pp. 1399–1409.

Ramankutty, N., and Foley, J., 1998, Characterizing patterns of global land use: An analysis of global croplands data. *Global Biogeochemical Cycles*, **12**, pp. 667–685.

See, L.M. and Fritz, S., 2005, A user-defined fuzzy logic approach to comparing global land cover products. In *14th European Colloquium on Theoretical and Quantitative Geography*, 9–12 September 2005, Lisbon, Portugal.

Siska, P.P. and Hung, I.-K., 2001, Assessment of kriging accuracy in the GIS environment. In *The 21st Annual ESRI International User Conference*, 9–13 July 2001, San Diego, CA.

Torbick, N., Qi, J., Ge, J., Olsen, J. and Lusch, D., 2005, An assessment of AFRICOVER and GLC2000 using general agreement and videography techniques. *IEEE International Proceedings: Geoscience and Remote Sensing Symposium, IGARSS'05*, **7**, pp. 5005–5008.

UN Millennium Project, 2005, Investing in development: a practical plan to achieve the Millennium Development Goals. Available at: http://www.unmillenniumproject.org/reports/index_overview.htm (accessed January 2008).