Understanding agricultural drivers of deforestation through remote sensing

Opportunities and limitations in sub-Saharan Africa

Samuel Bowers, Casey Ryan and Xiaoting Hou Jones

Working Paper
January 2017

Forests; Food and agriculture

Keywords:
Food, forestry, sub-Saharan Africa, deforestation, remote sensing

iied

THE UNIVERSITY *of* EDINBURGH

About the authors

Samuel Bowers is a postdoctoral researcher at The University of Edinburgh.

Casey Ryan is a senior lecturer at The University of Edinburgh.

Xiaoting Hou Jones is a researcher in the biodiversity team at IIED.

Corresponding author: Xiaoting.Hou.Jones@iied.org

Acknowledgements

Data acquisition and analysis was aided by the preliminary work of Dr Iain McNicol (The University of Edinburgh), who's advice was invaluable.

Partner organisation

This report was produced in partnership with the University of Edinburgh.

Produced by IIED's Natural Resources Group

The aim of the Natural Resources Group is to build partnerships, capacity and wise decision-making for fair and sustainable use of natural resources. Our priority in pursuing this purpose is on local control and management of natural resources and other ecosystems.

Published by IIED, January 2017

Samuel Bowers, Casey Ryan and Xiaoting Hou Jones (2017) Understanding agricultural drivers of deforestation through remote sensing: opportunities and limitations in sub-Saharan Africa. IIED Working Paper. IIED, London.

http://pubs.iied.org/17593IIED

ISBN 978-1-78431-446-0

International Institute for Environment and Development
80-86 Gray's Inn Road, London WC1X 8NH, UK
Tel: +44 (0)20 3463 7399
Fax: +44 (0)20 3514 9055
email: info@iied.org
www.iied.org

@iied
www.facebook.com/theIIED

Download more publications at www.iied.org/pubs

Effective monitoring of deforestation and cropland expansion in Africa requires reliable estimates of land cover area. However, continental scale land cover datasets generated solely or partially through remote sensing technologies show large differences in the extent and spatial distribution of forest and cropland. In this working paper we identify 14 of the most commonly used land cover data products, summarise their spatial, temporal and thematic properties, and compare their estimates of forest and cropland area. We describe the limitations to land cover data that result from their divergent predictions and discuss the implications for using this data to assess the agricultural drivers of deforestation in Africa.

Contents

Executive summary 4

1 Introduction 6
1.1 Background 6
1.2 Aims 6
1.3 The importance of land cover data 7
1.4 Sources of land cover data 7
1.5 Properties of land cover data 8

2 Methods 10
2.1 Data acquisition and processing 10
2.2 Data analysis 11
2.3 Study regions 12

3 Results 13
3.1 Forest and cropland area estimates 13
3.2 Spatial variation in forest and cropland area estimates 17
3.3 Forest and cropland area change 22

4 Discussion 27
4.1 Sources of disagreement between data products 28
4.2 The utility of land cover data for land use planning 30
4.3 Future developments in land cover data 31

5 Concluding remarks 33

Appendix 1: Documentation for assessed data products 34

Acronyms 52

References 53

Related reading 55

Executive summary

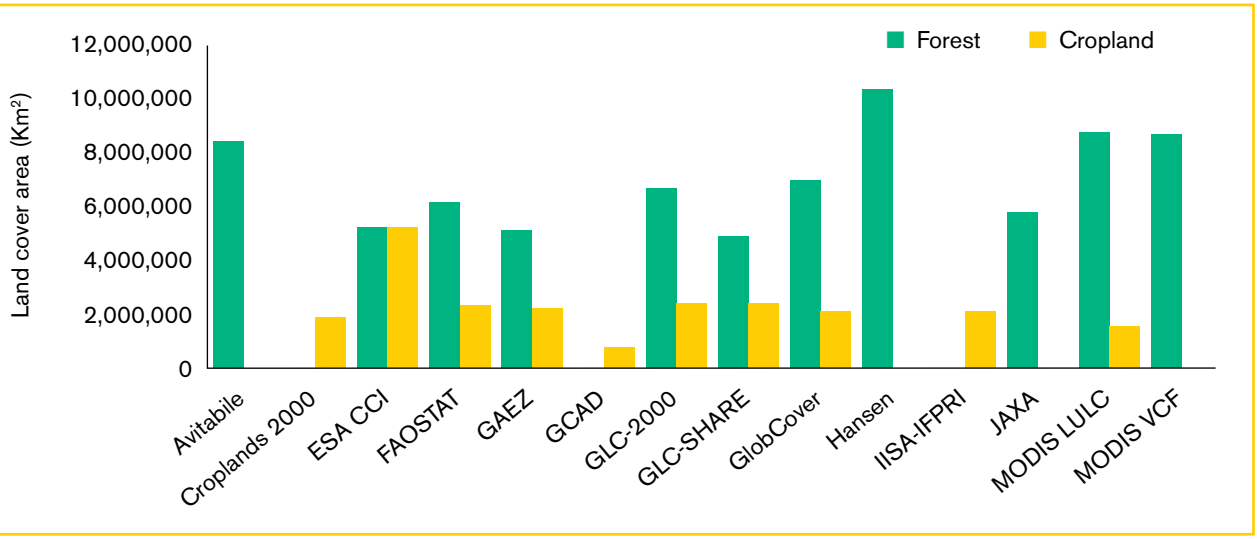
Reliable land cover data are important for monitoring the impacts of deforestation and cropland expansion in Africa, but uncertainties within such data are considerable. We examine 14 of the most commonly used land cover data products, generated solely or partially through remote sensing technologies, summarise their properties, and analyse the area of forest and cropland predicted by each dataset.

We present estimates of forest and cropland area and land cover change from a range of land cover data products at the continental spatial scale for sub-Saharan Africa, the national scale for Ethiopia, Ghana and Tanzania, and the regional scale for the Oromia region of Ethiopia, western region of Ghana, and Lindi region of Tanzania. We also identify some of the main sources of disagreement between data products and discuss the consequences of these limitations on measuring rates of forest loss.

Key findings are:

- Forest and cropland area estimates vary widely between data products (Figure 1).
- Disagreement between data products is apparent at continental, national and regional scales, being particularly elevated at the smallest spatial scales.
- Forest area estimates agree most strongly in areas of dense forest and where forest is clearly not present (eg deserts). Agreement between estimates of forest area is poor in areas of woodland and savannah.
- Crop area estimates show greater disagreement than for forest area, and may be particularly uncertain where agriculture is small scale and spatially scattered as is dominant in much of Africa. Data products agree better on locations where agriculture is limited rather than widespread.

Figure 1. Estimates of forest and cropland area in sub-Saharan Africa show widespread disagreement between data products, implying large uncertainties to the measurement of land cover in Africa.



[1] From here on shortened to *Muungano*.

- Most data products agree that Africa is undergoing a period of forest area loss and cropland area expansion, though estimates of the magnitude of these changes do not yet appear to be robust.
- Data products disagree in part due to their varied definitions of forest and cropland, which can be difficult to harmonise.
- Classification error is likely to be significant, resulting in error that is greater in magnitude than that of rates of land cover change.

Overall, it is difficult to reliably answer many important questions relating to the agricultural drivers of deforestation using these data products. There are ways to overcome some of the limitations presented – for example, data users can carefully examine land cover definitions and choose those data products that serve their objectives best – and over the coming decade advances in satellite technologies will continue to improve the data available. However, for now, these land cover data sources only offer limited answers, and it is important to use the spatial information they generate with caution.

1 Introduction

1.1 Background

Research by IIED's forest and biodiversity teams reveals that there is limited scope for meeting tripling domestic food demand in sub-Saharan Africa (SSA) by increasing food imports, reducing food waste and increasing crop yields. As a result, agriculture will continue to expand into forested areas to an extent that exceeds the current forest conservation targets of many countries.

The competition and trade-offs between the Sustainable Development Goals of ending hunger and conserving forests need to be recognised, better understood and, where necessary, proactively negotiated. With the support of country partners, IIED is exploring the political economy behind these trade-offs to shed light on the opportunities for achieving more joined-up land use policy and planning. This research is focused on Ethiopia, Tanzania and Ghana.

IIED is working in partnership with the University of Edinburgh to examine the strengths and limitations of using various land cover datasets generated solely or partially through remote sensing technologies to monitor the impacts of cropland expansion on forests. This information will enable a more accurate assessment of deforestation drivers (eg export commodities versus commodities for domestic consumption) and inform policy discussions at national and international level.

1.2 Aims

Monitoring of forest loss and identification of its causes requires reliable data on land cover and land cover change. A range of data products exist that provide estimates of forest and cropland area, but they are known to have large uncertainties and biases that are regionally variable. In this working paper we identify available sources of land cover data, compare their estimates of forest and cropland area, and identify areas of particular uncertainty in understanding the nature of cropland expansion and forest loss in Africa.

Specifically we:

- Identify existing high-quality data products that estimate crop and forest area and compare their accuracy as reported in the literature, their thematic and spatial resolution, and their definition of the targeted land cover.
- Present the results of our analysis of 14 data products. Our analysis includes a multiscale inter-comparison of the area estimates from each product, for crop and forest area separately. Scales analysed were continental (SSA), national (Ethiopia, Ghana and Tanzania) and regional (Oromia region of Ethiopia, western region of Ghana, Lindi region of Tanzania). The multiple scales are important because the products are used at different scales for different purposes and errors are likely to cancel out at large scales but may be significant at regional scales.
- From the evaluation of our results we draw lessons about the level of uncertainty in land cover estimates and the implications for our understanding of the relationship between agriculture and forest loss.

1.3 The importance of land cover data

The Food and Agriculture Organisation (FAO) of the United Nations defines land cover as "the observed [bio]physical cover of the Earth's surface". It is a term distinct from land use, which refers to the "arrangements, activities and inputs people undertake in a certain land cover type to produce, change, or maintain it" (Di Gregorio and Jansen 1998). Examples of widespread global land cover types are grasslands, shrublands, forests and croplands.

The need for accurate measurement of the area, spatial distribution and temporal changes to the Earth's land cover is well recognised (Herold *et al.* 2012). Land cover plays a key role in the biogeochemistry of the Earth system as an important driver of global climate change. Reliable land cover data are important for land use planning, with information required by policy makers to inform and monitor the outcome of management interventions. Whilst the importance of reliable information on global land cover is understood, present data are subject to uncertainties and biases that can hinder their use in many applications (Fritz *et al.* 2011).

Measuring and monitoring land cover change is a particular challenge on the African continent, where ecological understanding is poor and observational networks are limited. Tropical Africa is experiencing rapid rates of forest loss, in large part the consequence of forest clearance for croplands (Geist and Lambin 2002) thought to result from the expansion of smallholder agriculture for the production of food and cash crops (Fisher 2010). Cropland area in sub-Saharan Africa is projected to increase by 50 per cent by 2050, though this estimate is associated with considerable uncertainty (Prestele *et al.* 2016).

1.4 Sources of land cover data

The production of globally reliable databases of land cover is difficult because of the large scales involved, limitations with mapping technology and the wide range of land cover types – ranging from deserts to rainforests – that must be taken into account. Moreover, there are unresolved challenges in maintaining consistent and useful definitions of important land cover types, with differences in definitions of forest and croplands resulting in non-trivial divergences in land cover classifications (FAO 2016).

A number of methods exist to generate estimates of land cover area, most commonly based on the use of remote sensing data from Earth observation satellites, but also on census and survey data which are frequently used to measure agricultural land cover. These methods can be broadly categorised into: optical remote sensing, radar remote sensing, census based, and hybrid approaches.

Optical remote sensing

Optical remote sensing measures the visible, near-infrared and shortwave-infrared solar radiation reflected from the land surface to form images of the Earth's surface. Land surface types can be distinguished by their spectral reflectance signatures, making it possible to infer land cover properties. For example, forests are green and reflect a large proportion of near-infrared radiation, and different croplands may have unique timing of leaf display (phenology) which influences their reflectance pattern over time. Divergent reflectance properties are used to classify pixels into land cover types. Optical remote sensing data have the advantage of being abundant, with data available from a large number of satellite sensors spanning several decades. However, optical remote sensing techniques can struggle to distinguish some reflectance signatures – for example, separating the phenology of trees from grasses, and croplands from grasslands can be problematic. Optical remote sensing is not robust to persistent cloud cover which can be a particular issue in the wet tropics.

Radar remote sensing

Radio detection and ranging (radar) satellites emit radio energy towards the Earth's surface and measure the energy reflected back to the satellite sensor (backscatter). Radar backscatter can be used to distinguish between forested and non-forested area because backscatter of long wavelength radiation is greater from trees than from bare land areas. Radar remote sensing is robust to changes in cloud cover and plant phenology, though is at present little used for the identification of croplands over large scales.

Census-based

National forest and crop area data are commonly collected by countries and international agencies. Data are based on national inventories and questionnaires. Census information often includes details on land use in addition to land cover, such as forest management practises, crop types and yields – details which are difficult or impossible to gather using remote sensing. These data are not usually explicitly spatially defined, being limited to regional or national estimates and to administrative units. Census data can be difficult to compare with remote sensing data as the result of use of a range of measurement protocols, and are usually focused on areas of economic interest (ie forestry and agricultural production) so may omit data from natural or semi-natural ecosystems.

Hybrid data

Recent sources of land cover data have sought to integrate multiple forms of data into a single consistent data product. For example, the GLC-SHARE land cover product uses the 'best available' high resolution regional, national and global land cover databases for a given area. However, the harmonisation of multiple databases into a single coherent set of land cover classes can be challenging, as discussed later.

1.5 Properties of land cover data

Datasets that describe land cover are usually encoded as a raster image – a data format consisting of a regular grid of pixels comparable to that used for digital photographs. In place of the colour information contained in the pixels of a photograph, pixels in land cover data contain data describing the land cover at that location. Geospatial raster images are stored in a range of specialised formats which contain metadata that make pixels spatially locatable and allow the quantitative analysis and comparison of land cover datasets. Geospatial raster images have a range of properties. For the quantification of forest and cropland area important dataset properties are: spatial resolution, temporal resolution and thematic resolution.

Spatial resolution

Spatial resolution refers to the ability to distinguish between (or resolve) two objects in an image. Image resolution in remote sensing data is measured as a ground sampling distance which describes the distance separating image pixels at the Earth's surface. In the highest resolution remote sensing data this can be as small as half a metre, though land cover data are usually available at coarser spatial resolutions of 30 metres to 10 kilometers. This compares to census data which is usually only available at regional or national scales. Spatial resolution is sometimes expressed in units of arcminutes or seconds – an angular measurement describing the proportion of the Earth's circumference represented by a pixel. In this system the Earth's circumference is divided into 360 degrees, with each degree containing 60 arcminutes and each arcminute containing 60 arcseconds. The exact area represented in each pixel varies with latitude but at the equator 30 arcseconds is equivalent to approximately 1 kilometre and 5 arcminutes is equivalent to about 10 kilometres.

Temporal resolution

Temporal resolution refers to the land cover re-measurement time at a given location. Land cover data most frequently provide a single estimate of land cover valid for a specific year or range of years. Other datasets provide estimates of land cover over multiple years and allow the computation of rates of land cover change. Temporal resolution in these data products is commonly one year but the re-measurement period can be as long as five years. There is a trade-off between spatial and temporal resolution in remote sensing data because both high spatial and temporal resolution cannot usually be achieved. Census-based land cover data collection tends to be infrequent and the methods and definitions employed may be inconsistent between surveys over time.

Thematic resolution

Different sources of land cover data represent the properties of different land cover types. Thematic resolution refers to the type of data encoded in a raster image and the level of detail of the information it contains. Land cover data are usually represented by discrete land cover categories, where each pixel is identified as a particular land cover category (eg forest or cropland). The number and detail of these categories varies between data products. Other data give estimates of a particular biophysical property (eg above ground biomass (AGB) or tree cover) on a continuous measurement scale (rather than by category) therefore giving more precise but less wide-ranging information on land cover characteristics. Census-based data can provide much greater thematic resolution, for example data on forest type or crop type.

Table 1. Data products used to assess variation in forest and crop area estimates.

DATA PRODUCTS	FOREST	CROP	LAND COVER METRIC	SPATIAL RESOLUTION	YEAR(S) REPRESENTED	KEY REFERENCES
Optical remote sensing:						
Croplands 2000		✓	% cover	5 arcminutes (~10km)	2000	Ramankutty *et al.* (2008)
ESA CCI	✓	✓	Categorical	300m	2000, 2005 & 2010	Bontemps *et al.* (2013)
GLC-2000	✓	✓	Categorical	32 arcseconds (~1km)	2000	Mayaux *et al.* (2004)
GlobCover	✓	✓	Categorical	300m	2005 & 2009	Bontemps *et al.* (2011)
Hansen	✓		Tree cover	30m	2000–2012	Hansen *et al.* (2013)
MODIS LULC	✓	✓	Categorical	15 arcseconds (~500m)	2000–2012	Friedl *et al.* (2002)
MODIS VCF	✓	✓	Tree cover	15 arcseconds (~500m)	2000–2012	Hansen *et al.* (2003)
Radar remote sensing:						
JAXA	✓		Categorical	100m	2007–2010	Shimada *et al.* (2014)
Census based:						
FAOSTAT	✓	✓	Categorical	By country	2000–2012	Smith (2010)
Hybrid data:						
Avitabile	✓		Biomass	30 arcseconds (~1km)	2005	Avitabile *et al.* (2012)
GAEZ	✓	✓	% cover	5 arcminutes (~10km)	2000	Fischer *et al.* (2008)
GCAD		✓	Categorical	32 arcseconds (~1km)	Spatially variable	Teluguntla *et al.* (2015)
GLC-SHARE	✓	✓	% cover	30 arcseconds (~1km)	Spatially variable	Latham *et al.* (2014)
IIASA-IFPRI		✓	% cover	30 arcseconds (~1km)	Spatially variable	Fritz *et al.* (2015)

2 Methods

2.1 Data acquisition and processing

Land cover data were obtained from 14 datasets representing a range of data sources and land cover classification methods (Table 1). We limited our analysis to data from the year 2000 or later and to the most commonly used datasets. Of the selected data products, 11 provide estimates of forest area and 11 provide estimates of crop area. The properties of data products included in this analysis and their reported accuracy are detailed in Appendix 1.

Each selected data product was downloaded for all available years along with associated documentation describing the methods used to generate each dataset. Images were reprojected to an equal-area projection (Lambert cylindrical equal-area) with the geospatial data abstraction library. A pixel resolution of 309 metres was specified for all data sources equivalent to 1/360th of a degree (10 arcseconds) at the equator. Data were resampled using nearest-neighbour resampling, except where high resolution data allowed for a reliable within-pixel land cover proportion to be calculated (ie the 30 metre resolution Hansen dataset).

Land cover classes differ between land cover products. Some specify tree cover (eg Hansen, MODIS VCF), others forest cover (eg JAXA), others biomass (eg Avitabile), others proportional land cover by forests and croplands (eg GLC-SHARE) and, most commonly, others specify discrete land cover classes (eg ESA CCI, GlobCover). Where mixed classes exist in data products, we weighted its area by the proportion of that land cover class that relates to forest/cropland. An example is 'mosaic cropland' which may contain areas of both forest and cropland.

Complicating the inter-comparison of different data products, there are no universally agreed definitions of forest and cropland. We aim to use a consistent definition of forest and cropland to compare data products, though it should be noted that a fully-consistent inter-comparison is not always possible. Where possible we use the widely applied FAO definitions of forest and cropland which are:

Forest: Land with tree crown cover of more than 10 per cent over an area of more than 0.5 hectares. This is assumed to occur where biomass is greater than 10 tonnes of carbon per hectare, a figure consistent with data from southern and eastern African woodlands. Forest definitions in classified land cover products vary, and are not always well documented; we aim to include each land cover class that is likely to meet the FAO definition.

Cropland: Areas of annual (eg maize) or perennial (eg coffee) crops. Forest plantations and pasture are not included in this definition.

2.2 Data analysis

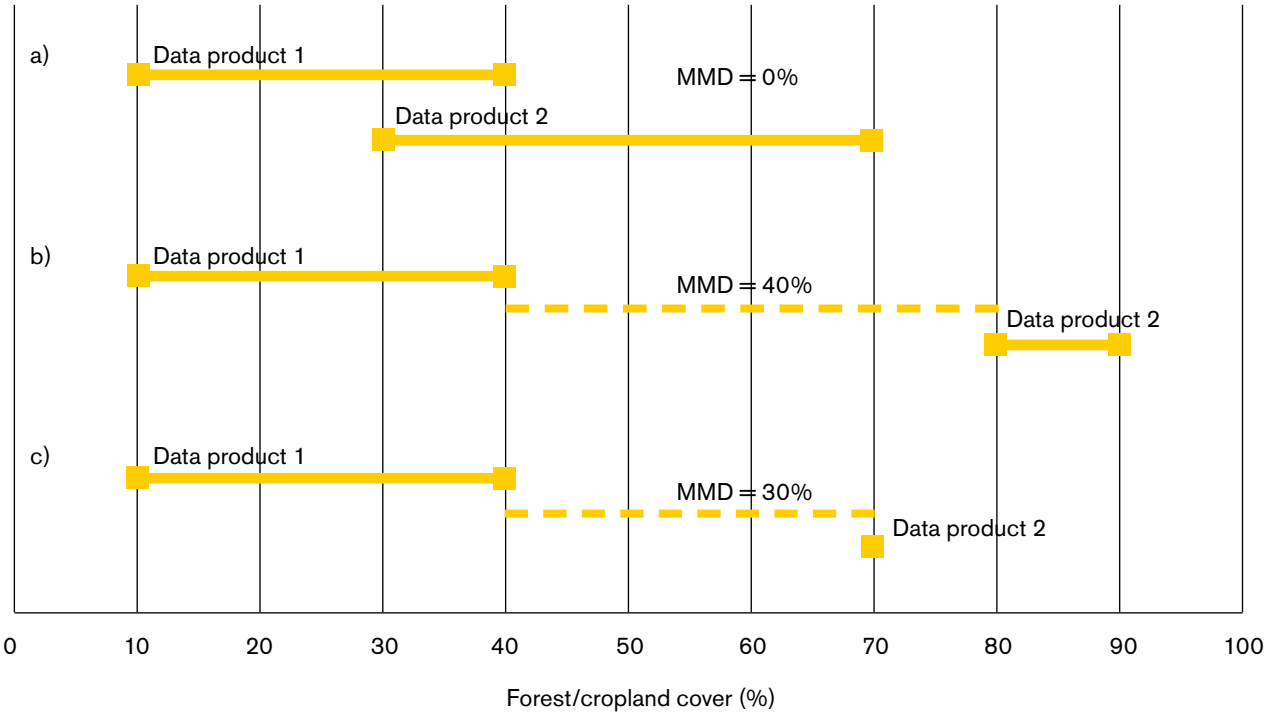
The area of forest and cropland estimated by each data product was calculated by multiplying the proportional land cover of each pixel by the pixel area (0.0995 square kilometres). A real coverage of forest and cropland were compared using data from the most recently available year from each data product. There will be some error introduced in comparing forest/cropland in 2000 (eg GLC-200) with 2012 (eg Hansen) though changes to land cover over this time period are expected to be small relative to intrinsic differences between data products. Where a class or category in a data product represents a range of possible forest/cropland cover we assume the central value to be the 'true' land cover proportion of that pixel. The 'coefficient of variation' (ratio of standard deviation to the mean) of land cover areas is used as a metric of how variable estimates of forest and cropland area are between data products.

Differences in forest and cropland area estimates translate into variation in the spatial extent of land cover types. We characterise differences in spatial distributions of forest and cropland using two graphical methods:

i) Maps of how commonly a location is classified as cropland or forest in each of the assessed data products. In this comparison land cover probability varies from 0 per cent (no data product predicts cropland or forest) to 100 per cent (all data products predict cropland or forest). Where this figure is 50 per cent, data products show a 50:50 split between forest/not forest or cropland/not cropland. Where only a fraction of a pixel belongs to one of these classes we assume this to be the probability of the entire pixel representing forest or cropland.

ii) Maps of minimum measurable disagreement (MMD), an index of the smallest land cover area by which two data products might disagree (Figure 2). This conservative index of disagreement, based on the work of Fritz *et al.* (2011), was calculated for all paired comparisons between forest and cropland data products. The mean disagreement for each pixel was calculated to identify those locations that land cover products least likely to agree.

Figure 2. Minimum measurable disagreement (MMD) describes the smallest area by which two data products might disagree.



In example a) the range of land cover of the two products overlap so MMD is assumed to be 0 per cent. In b) the maximum possible cropland/forest cover of data product 1 is 40 per cent and minimum forest/cropland cover of data product 2 is 80 per cent, meaning that 40 per cent is the minimum degree to which these two products disagree. In c) data product 2 has only a single value, which we assume to have no error range.

Some data products provide estimates of forest and cropland area over multiple years (ESA CCI, FAOSTAT, GlobCover, Hansen, JAXA, MODIS LULC, MODIS VCF). These data products can be used to calculate rates of forest and cropland cover change. We assess variation in estimates of land cover change using two methods:

i) Tabulation of mean annual change in forest and cropland area, expressed as the mean average annual per cent change in forest and cropland area over the time period expressed by the dataset. This comparison assumes that rates of forest and cropland area change are constant through time.

ii) Graphically based on cumulative change in land cover area since the first measured year. As each data product predicts widely different initial forest and cropland area and represents a different time period we express land cover change as a rate relative to the first measured year. An increase of forest/cropland above 100 per cent corresponds to a net gain in that land cover type, and a fall below 100 per cent represents a net loss of that land cover type.

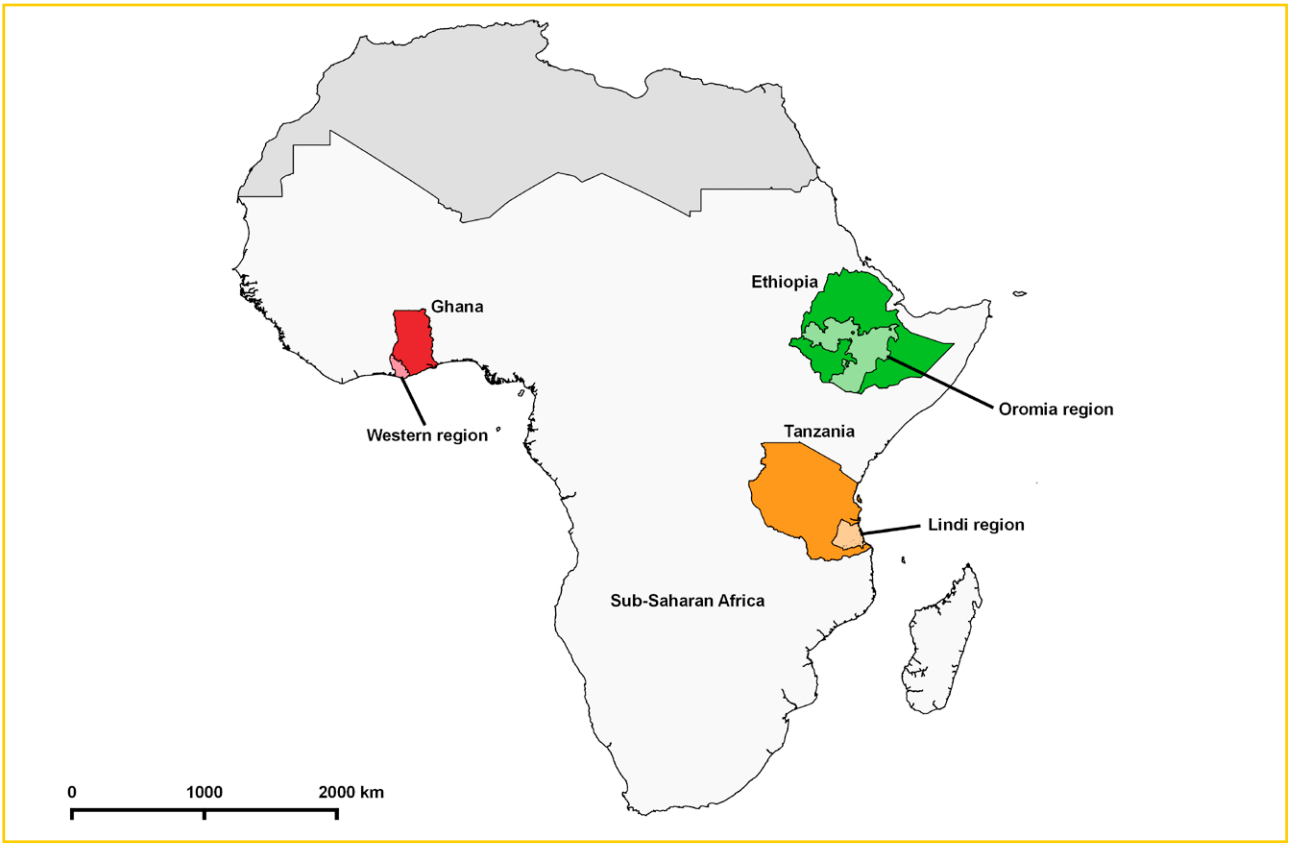
2.3 Study regions

Aerial land coverage by forests and croplands was calculated for each data product at the continental spatial scale of sub-Saharan Africa (defined as the African continent excepting Algeria, Egypt, Libya, Morocco, Tunisia, and Western Sahara), at the national scale of Ethiopia, Ghana, and Tanzania, and at the regional scale of the Oromia region of Ethiopia, the western region of Ghana, and the Lindi region of Tanzania (Figure 3). These all represent locations that have experienced the recent conversion of forests into agricultural land. The total land area of each of these study regions is presented in Table 2.

Table 2. Total area of selected study regions, including coverage by both land and water.

STUDY REGION	TOTAL AREA (KM²)
sub-Saharan Africa	24,402,064
Ethiopia	1,142,034
Ghana	241,685
Lindi region of Tanzania	950,256
Oromia region of Ethiopia	327,929
Western region of Ghana	24,707
Lindi region	65,662

Figure 3. Location of seven selected study regions classed as continental (sub-Saharan Africa), national (Ethiopia, Ghana and Tanzania) and regional (Oromia region of Ethiopia, western region of Ghana, Lindi region of Tanzania) spatial scales.



3 Results

3.1 Forest and cropland area estimates

Forest area estimates differ substantially between data products (Table 3 and Figure 4). For instance, estimates of forest area for SSA range from 4,930,000 km² (GLC-SHARE) to 10,370,000 km² (Hansen). Cropland area estimates are also widely variable with cropland area in sub-Saharan Africa ranging from 830,000 square kilometres (GCAD) to 5,180,000 square kilometres (ESA CCI), a six fold difference in cropland area estimates between data products. Coefficients of variation suggest that estimates of forest area are more uncertain at national and regional scales than for the entire continent. Variation in cropland area estimates is high at all of the scales considered with uncertainty in cropland detection particularly pronounced in Tanzania.

Differences in forest and cropland area estimates at continental, national and regional scales are presented graphically in Figures 4–10.

Figure 4. Forest and cropland area predictions for land cover products in sub-Saharan Africa.

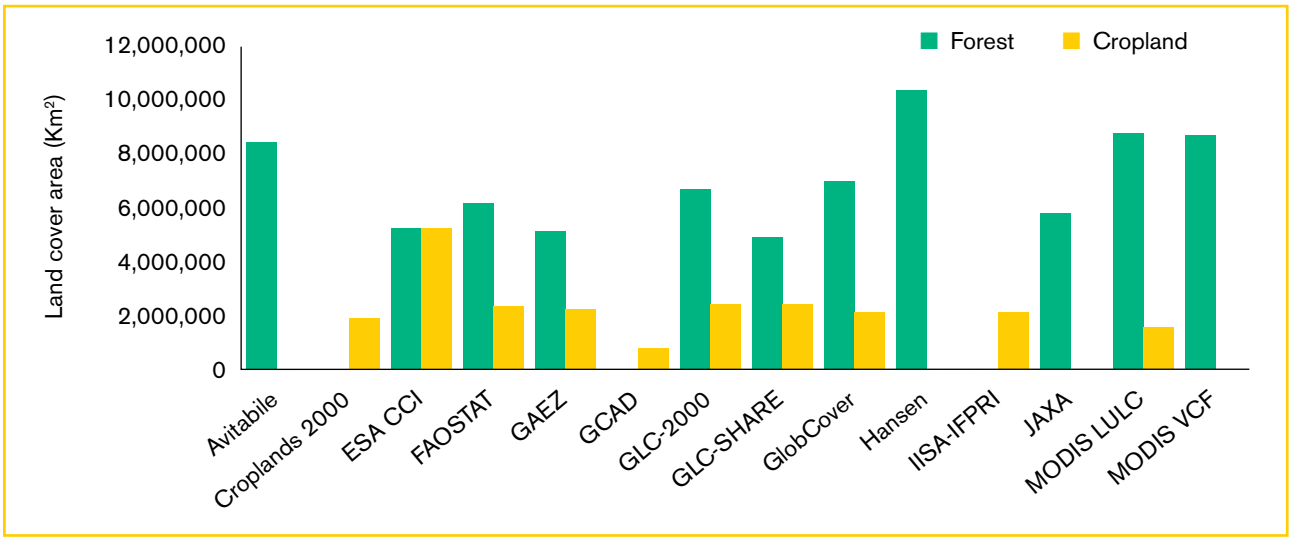


Table 3. Forest and cropland area estimates for data products for sub-Saharan Africa, and at national and regional study locations. Units of area are km^2

	SUB-SAHARAN AFRICA		ETHIOPIA		GHANA		TANZANIA		OROMIA REGION		WESTERN REGION		LINDI REGION	
Data product	Forest	Cropland	Forest	Crop-land	Forest	Crop-land	Forest	Crop-land	Forest	Crop-land	Forest	Crop-land	Forest	Crop-land
Avitabile	8,436,541	–	398,636	–	113,637	–	433,602	–	137,334	–	23,385	–	49,642	–
Croplands 2000	–	1,937,264	–	107,745	–	60,616	–	51,347	–	36,852	–	3,295	–	2,238
ESA CCI	5,231,549	5,180,339	83,086	359,337	53,862	104,169	259,404	360,777	36,847	150,359	6,238	14,423	27,091	22,221
FAOSTAT	6,201,673	2,364,065	124,178	162,590	92,802	74,000	468,040	–	–	–	–	–	–	–
GAEZ	5,120,566	2,242,729	84,561	149,227	60,468	61,470	245,714	120,580	33,952	57,997	11,253	7,672	26,631	4,082
GCAD	–	833,099	–	65,931	–	21,485	–	38,237	–	25,352	–	2,453	–	1,208
GLC-2000	6,726,626	2,435,996	82,943	311,297	122,872	29,919	281,257	169,720	25,755	125,417	14,958	6,126	31,840	4,945
GLC-SHARE	4,928,306	2,419,171	72,218	269,527	44,565	61,308	337,121	144,750	23,452	104,409	6,352	8,665	40,532	6,120
GlobCover	6,963,576	2,141,009	269,134	189,685	69,667	34,760	402,561	97,318	102,349	75,957	10,993	6,497	35,284	7,768
Hansen	10,373,806	–	308,755	–	177,633	–	630,643	–	129,312	–	22,443	–	61,479	–
IISA-IFPRI	–	2,094,604	–	135,294	–	68,191	–	105,495	–	56,928	–	9,503	–	4,697
JAXA	5,763,441	–	300,699	–	73,195	–	239,429	–	125,920	–	20,844	–	25,092	–
MODIS LULC	8,750,876	1,595,528	260,129	155,270	101,193	43,222	528,570	54,663	101,440	48,797	10,476	4,255	48,018	1,943
MODIS VCF	8,734,265	–	322,911	–	124,716	–	440,300	–	134,326	–	21,859	–	52,744	–
Mean average	7,021,021	2,324,380	209,750	190,590	94,056	55,914	387,876	126,988	85,069	75,785	14,880	6,988	39,835	6,136
Coefficient of variation	26%	49%	59%	50%	43%	45%	34%	79%	59%	57%	46%	54%	33%	107%

The coefficient of variation (standard deviation / mean) is a measure of relative variation between land cover area estimates. Where a data product gives estimates over multiple years, we take the most recent estimate of land cover.

Figure 5. Forest and cropland area predictions for land cover products in Ethiopia.

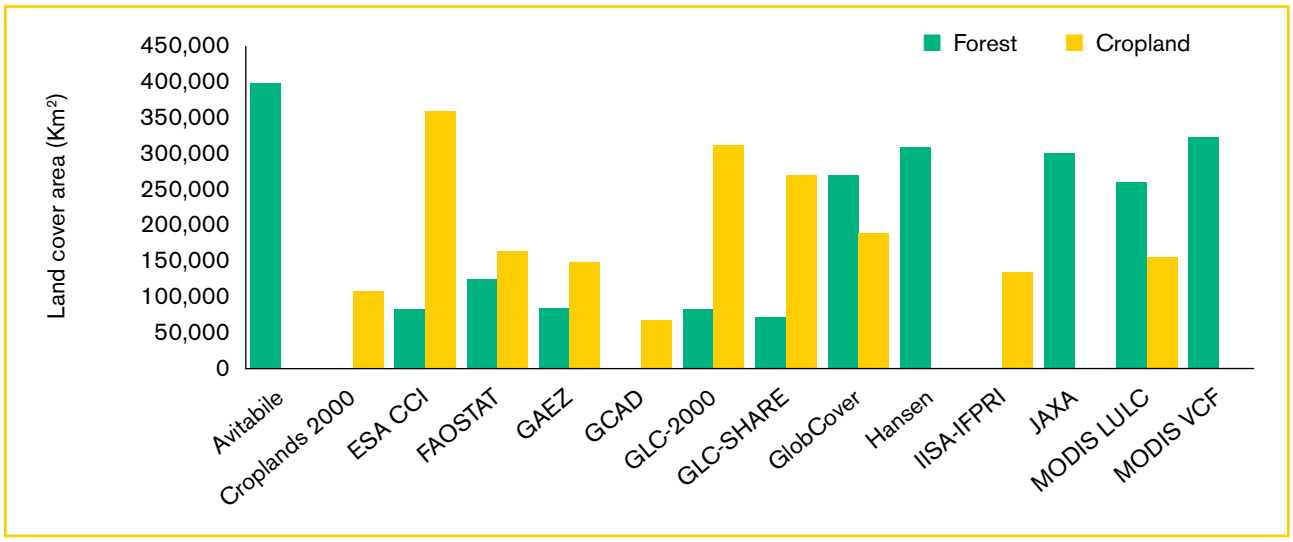


Figure 6. Forest and cropland area predictions for land cover products in Ghana.

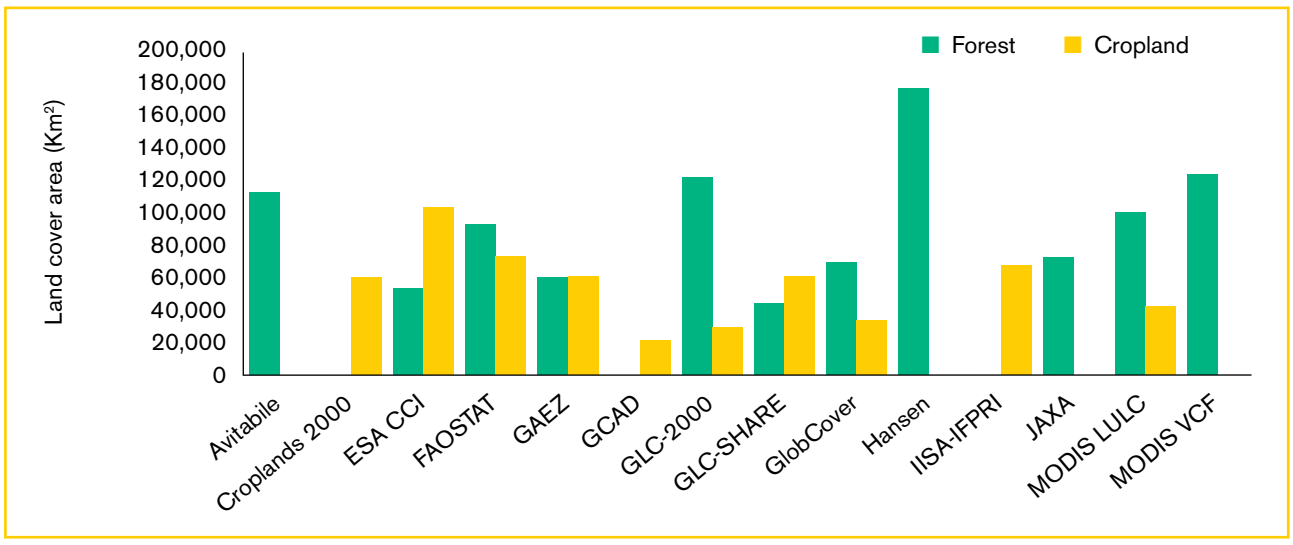


Figure 7. Forest and cropland area predictions for land cover products in Tanzania.

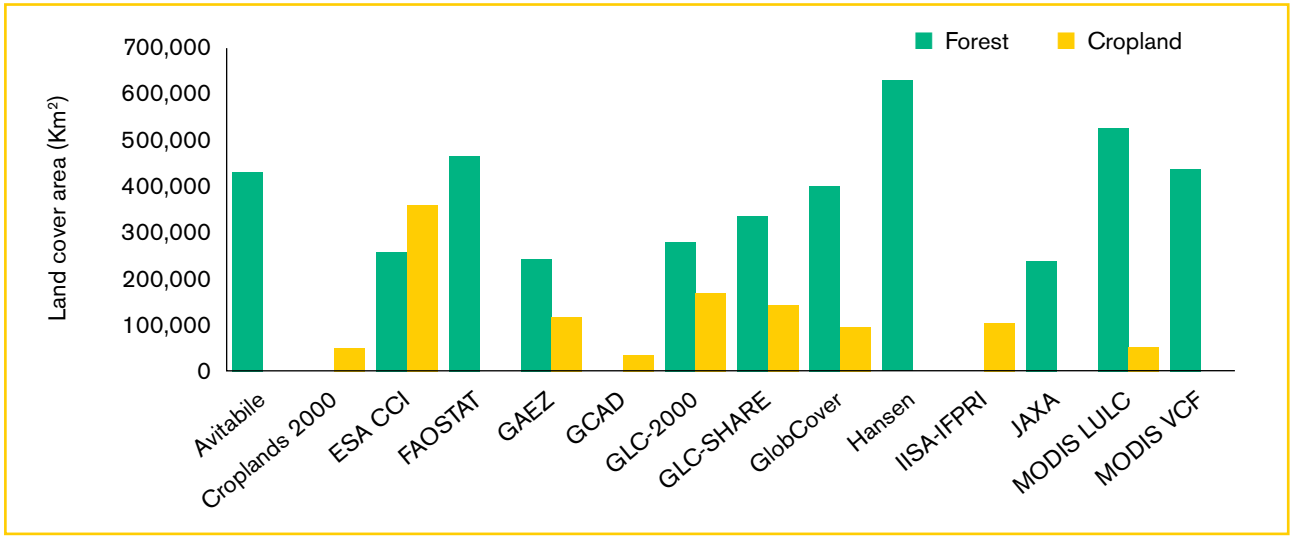


Figure 8. Forest and cropland area predictions for land cover products in Oromia region of Ethiopia.

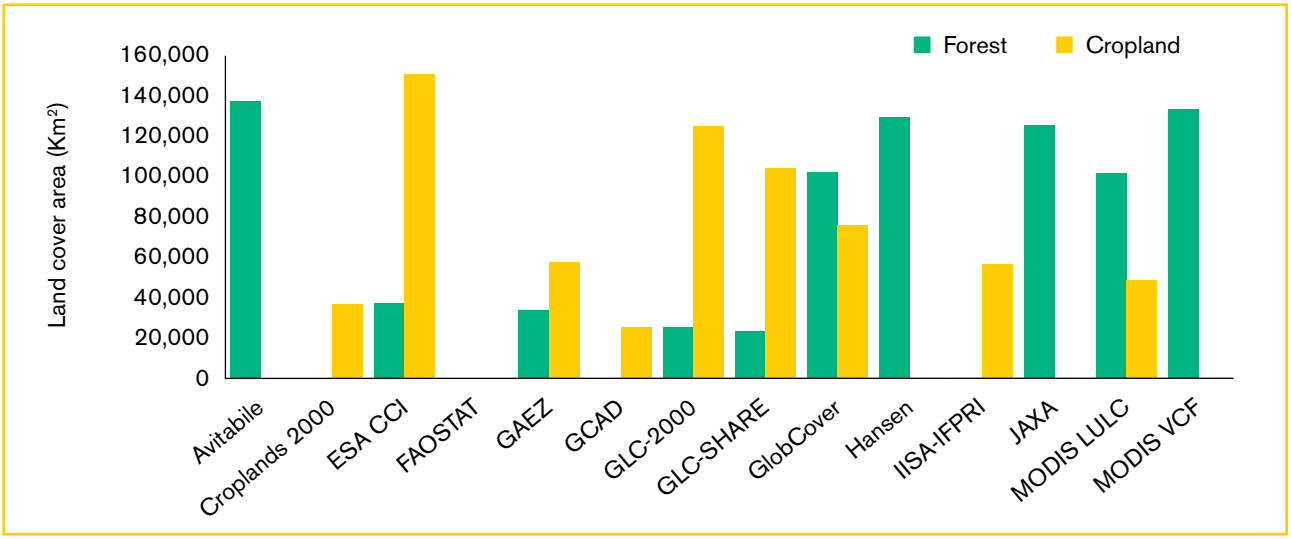


Figure 9. Forest and cropland area predictions for land cover products in the western region of Ghana.

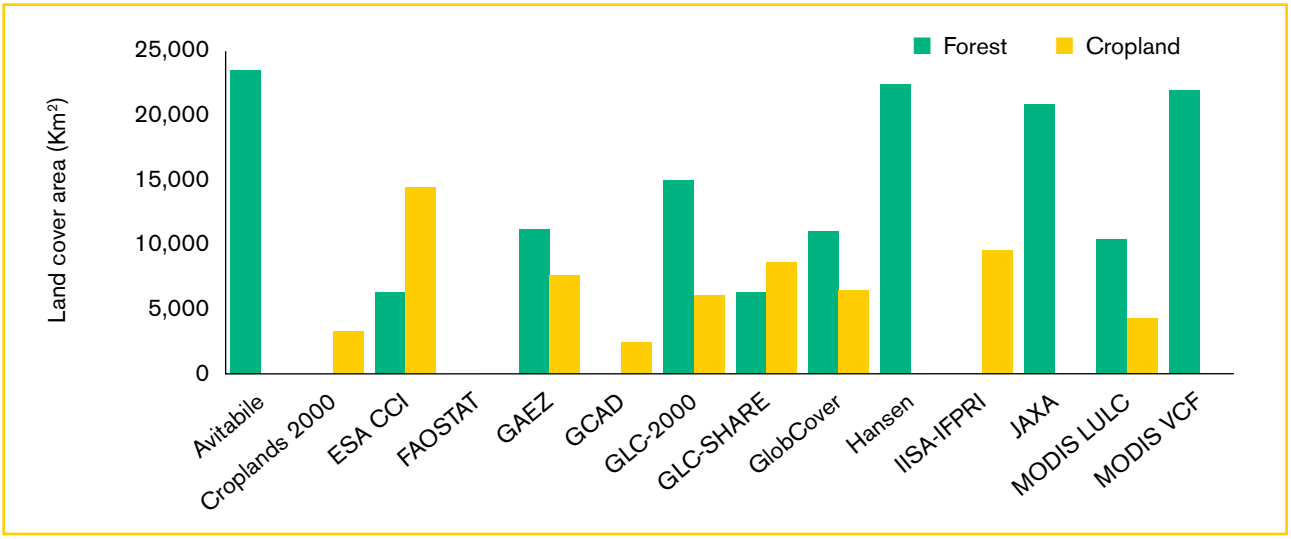
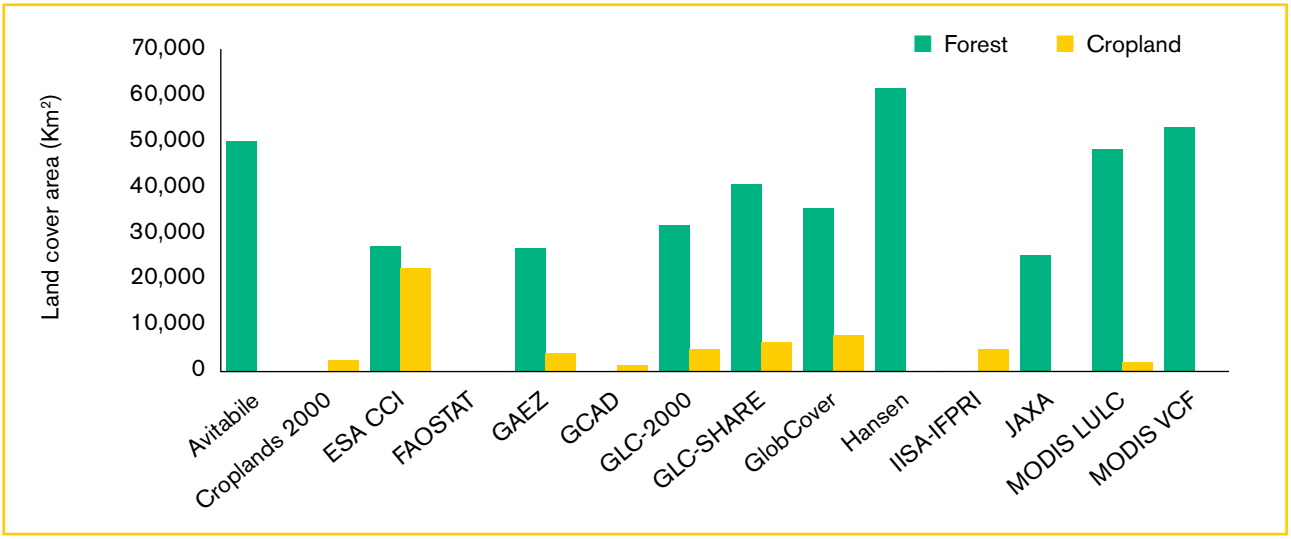


Figure 10. Forest and cropland area predictions for land cover products in Lindi region of Tanzania.



3.2 Spatial variation in forest and cropland area estimates

The differences we observe in forest and cropland area estimates necessarily translate into variation in the spatial extent of land cover types. Maps of the probability of a pixel being classed as forest or cropland are presented in Figures 11–17. Areas where there are differences in the forest and cropland area estimates of data products are highlighted by visualising average MMD (Figures 18–24). Locations with a high MMD indicate areas where disagreement and land cover uncertainty is greatest between data products.

In general, there is a good agreement amongst data products in areas where forests are dense (eg Congo Basin, Western Tanzania and Southern Ghana) and in areas where forests are clearly absent (eg Sahara desert, eastern Ethiopia, dense agriculture in southern Ghana). Poor agreement between data products is very apparent in the woodland and savannah ecosystems of southern Africa which data products inconsistently characterise as forest. This is especially clear in the Lindi region where there is very limited agreement in forest area between data products associated with widespread savannah woodlands.

Very few locations show 100 per cent agreement for the presence of croplands, with the most apparent convergence of cropland area estimates in the Sahel, central Ethiopia and in parts of Ghana. Disagreement between data products is mostly limited to areas of known dense agriculture with different data products largely agreeing on locations where croplands are absent or rare. There is noticeably little disagreement between data products in the western region of Ghana (Figure 16) where a comparatively simple mosaic of dense forest and intensely-used agricultural land results in a relatively straightforward land cover classification in all data products.

Figure 11. Probability of land cover products predicting the presence of forests and croplands in sub-Saharan Africa.

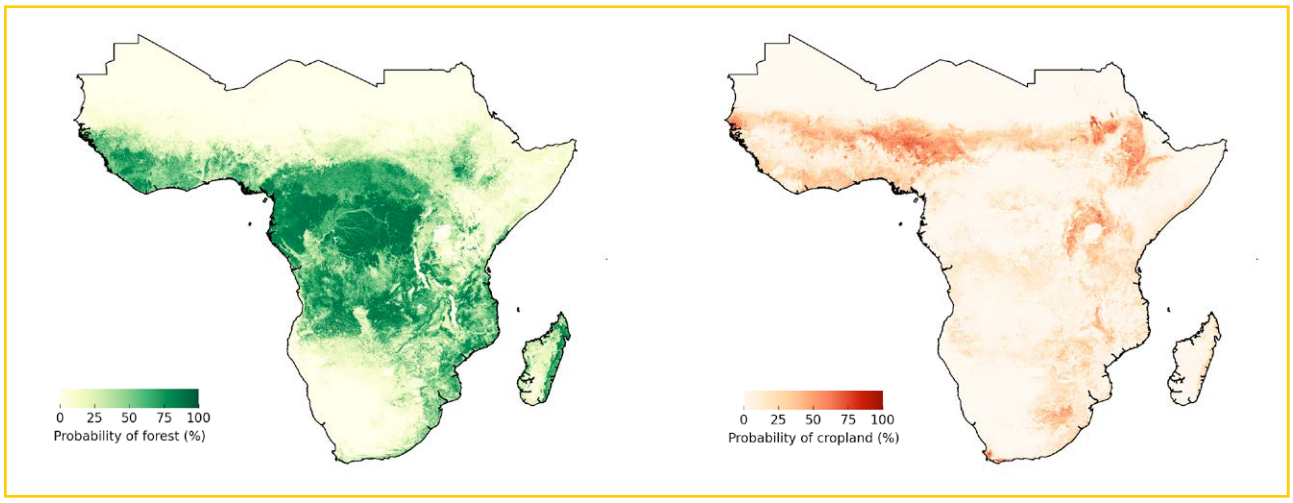


Figure 12. Probability of land cover products predicting the presence of forests and croplands in Ethiopia.

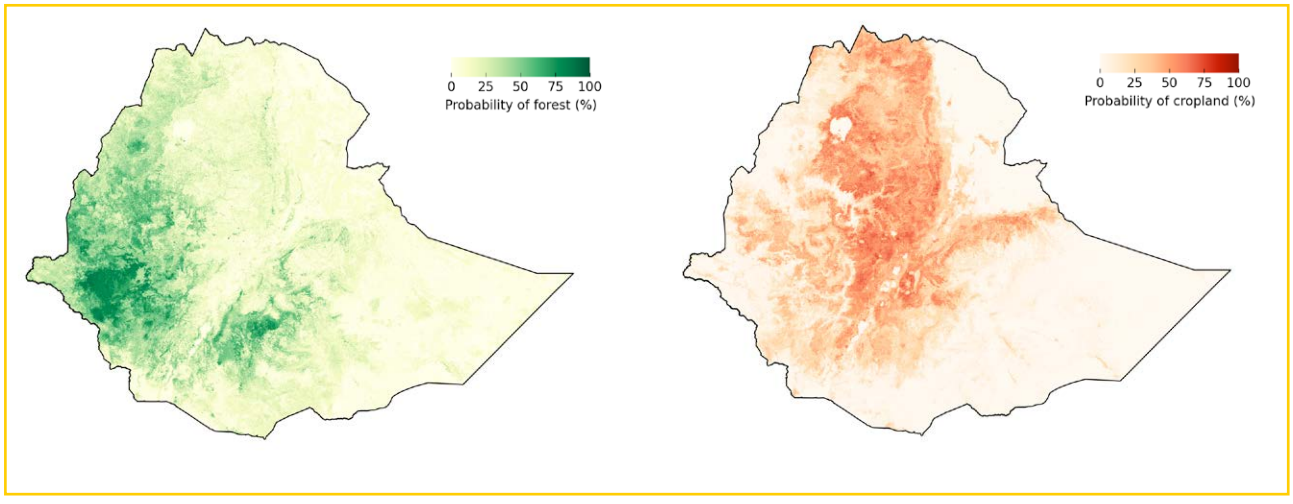


Figure 13. Probability of land cover products predicting the presence of forests and croplands in Ghana.

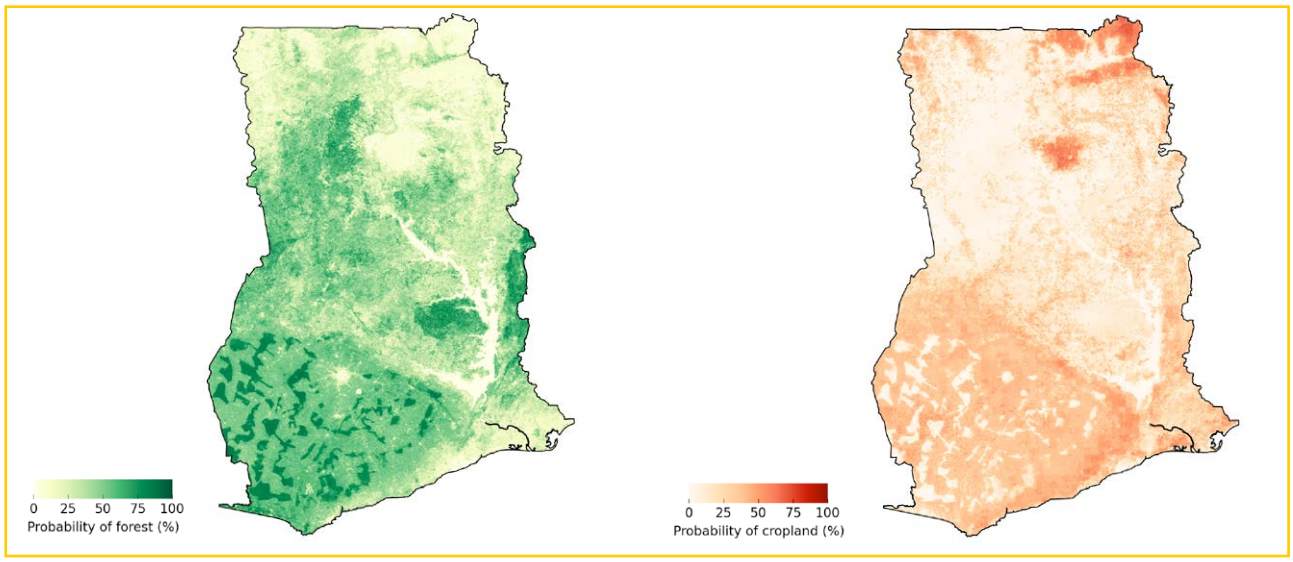


Figure 14. Probability of land cover products predicting the presence of forests and croplands in Tanzania.

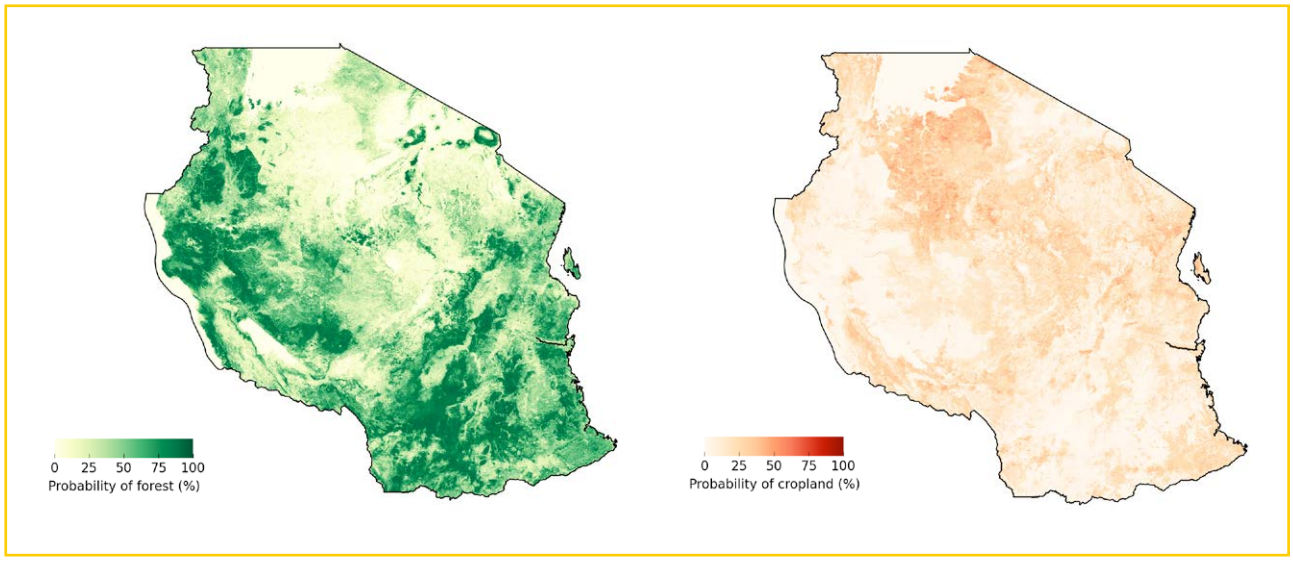


Figure 15. Probability of land cover products predicting the presence of forests and croplands in Oromia region.

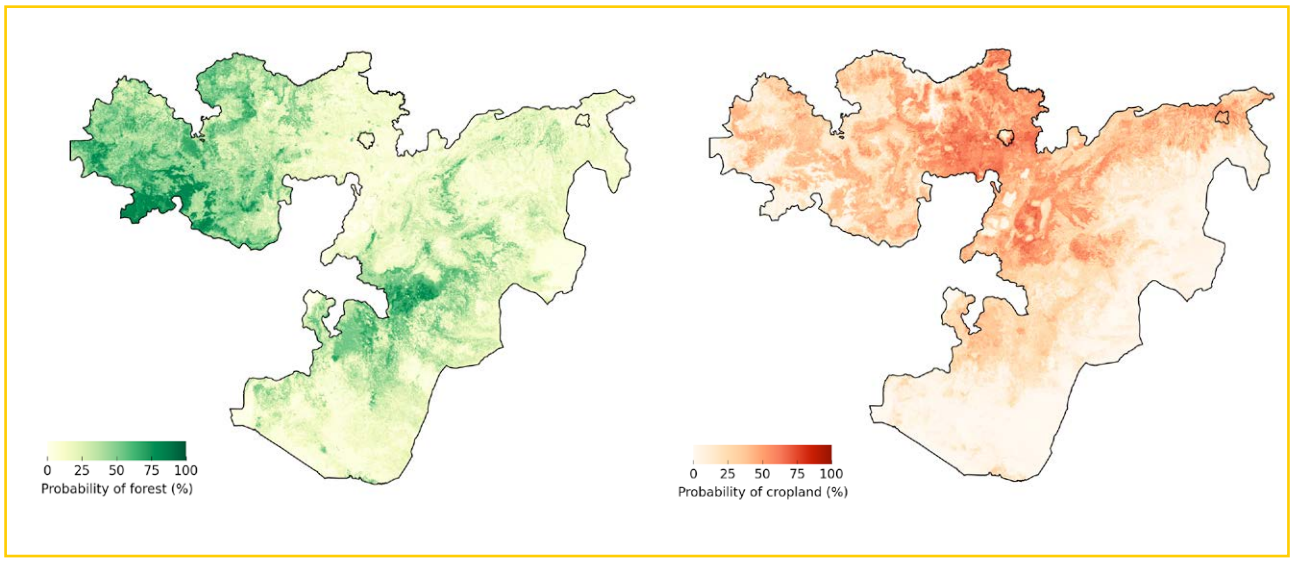


Figure 16. Probability of land cover products predicting the presence of forests and croplands in the western region of Ghana.

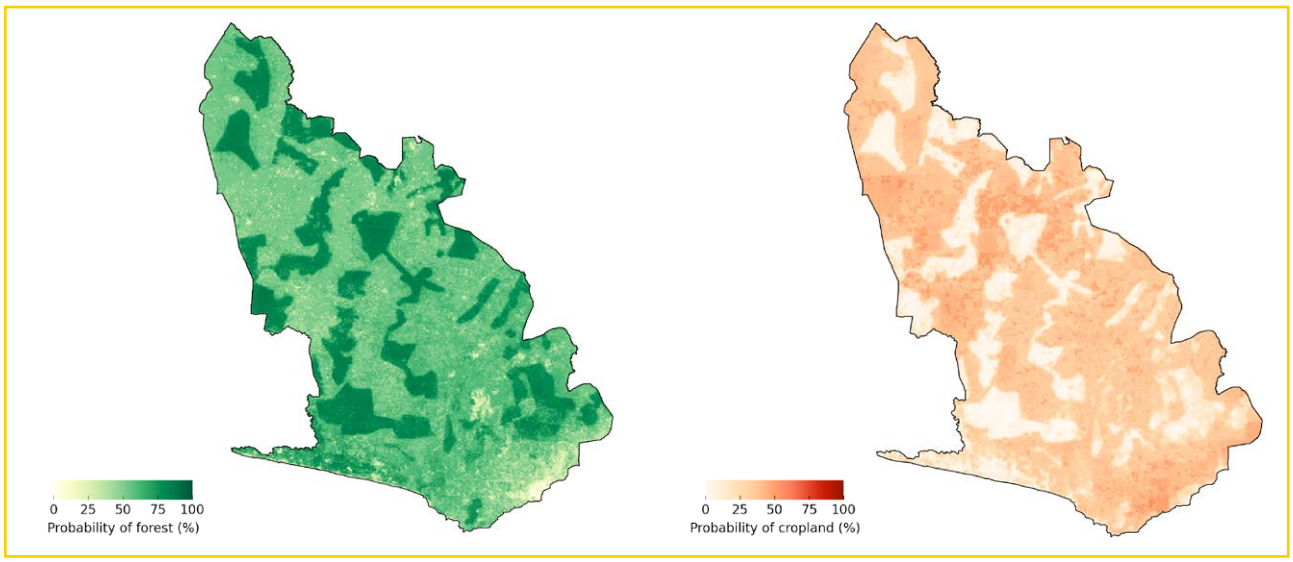


Figure 17. Probability of land cover products predicting the presence of forests and croplands in Lindi region.

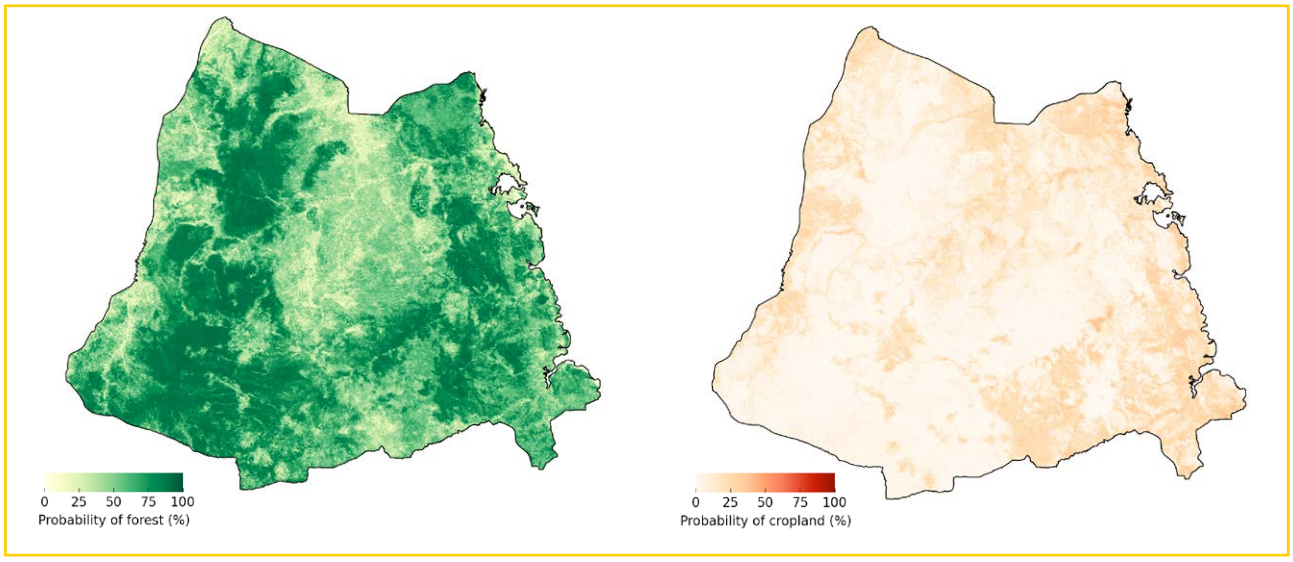


Figure 18. Average minimum measurable disagreement (MMD) in sub-Saharan Africa between forest and cropland area estimates in data products.

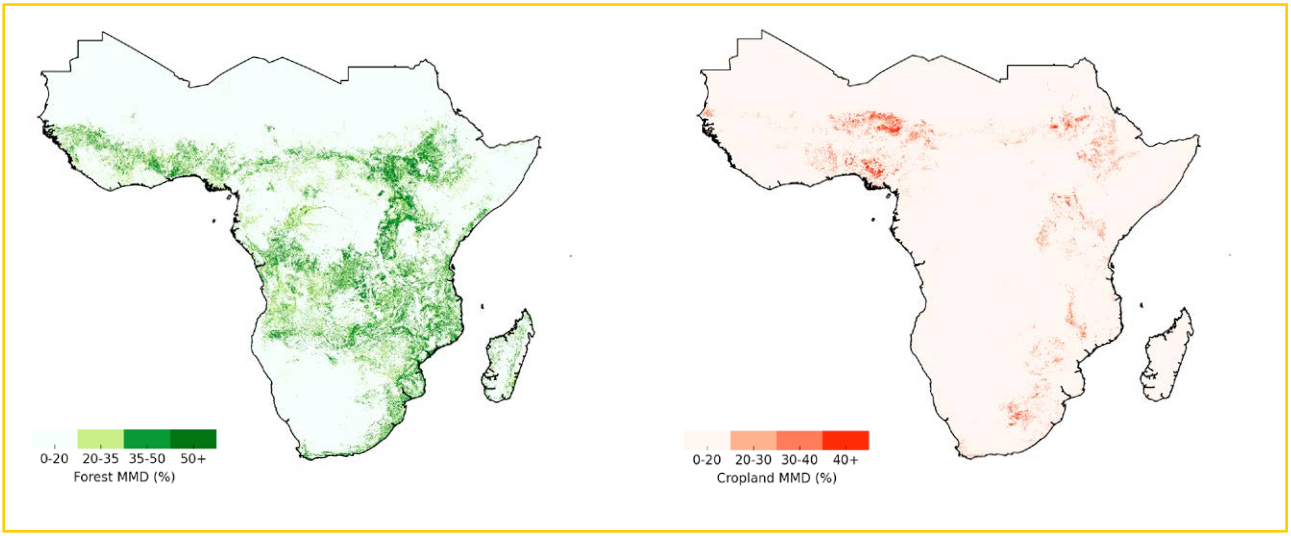


Figure 19. Average minimum measurable disagreement (MMD) in Ethiopia between forest and cropland area estimates in data products.

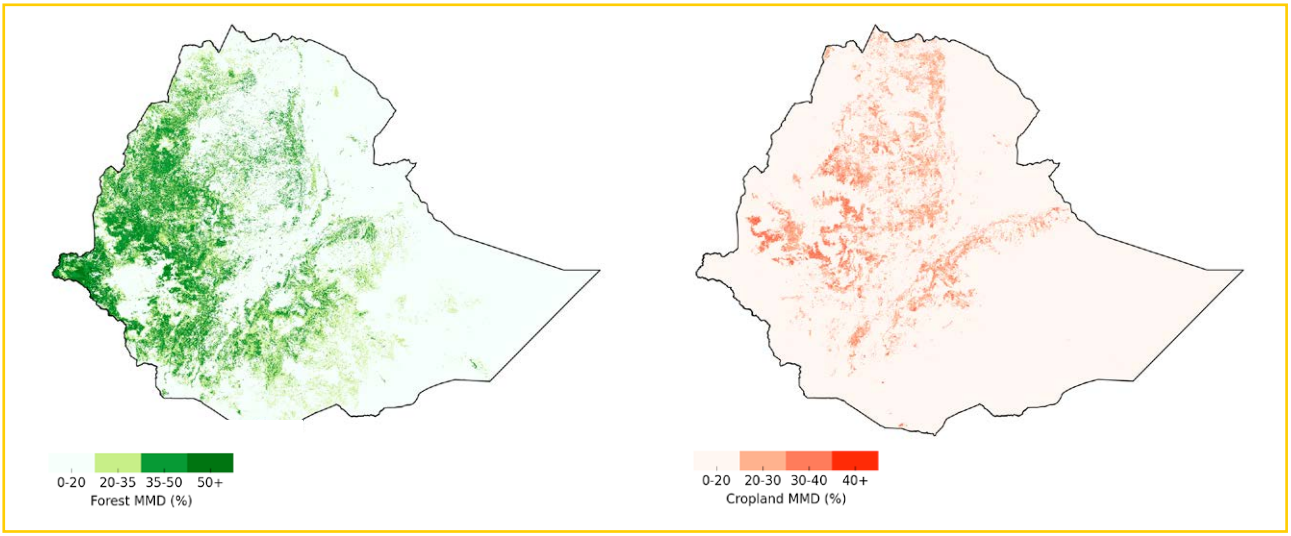


Figure 20. Average minimum measurable disagreement (MMD) in Ghana between forest and cropland area estimates in data products.

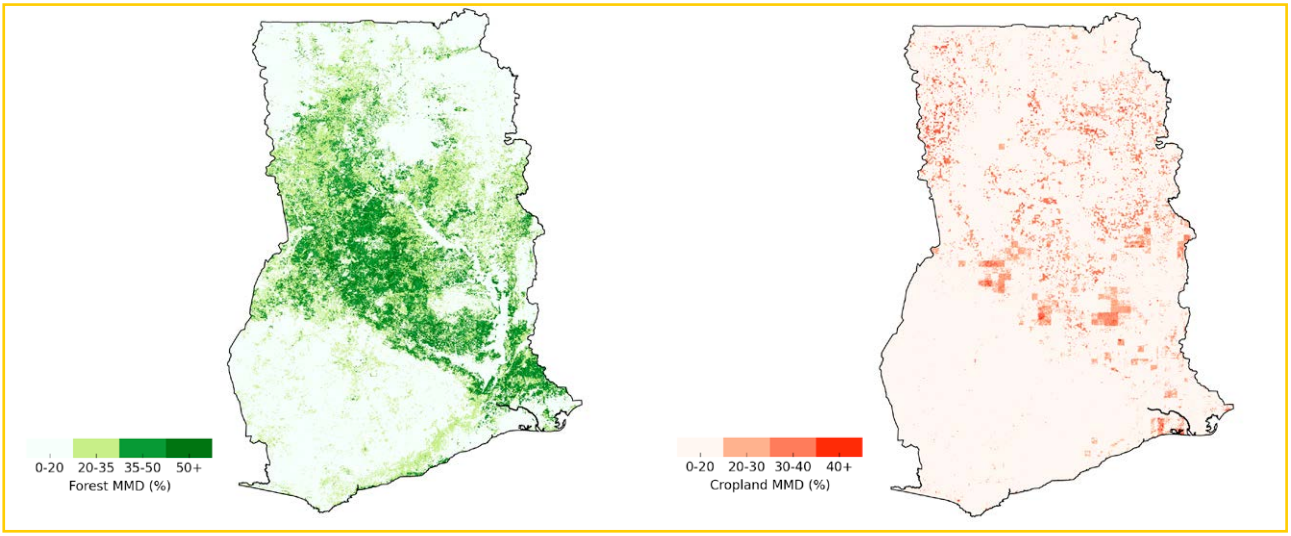


Figure 21. Average minimum measurable disagreement (MMD) in Tanzania between forest and cropland area estimates in data products.

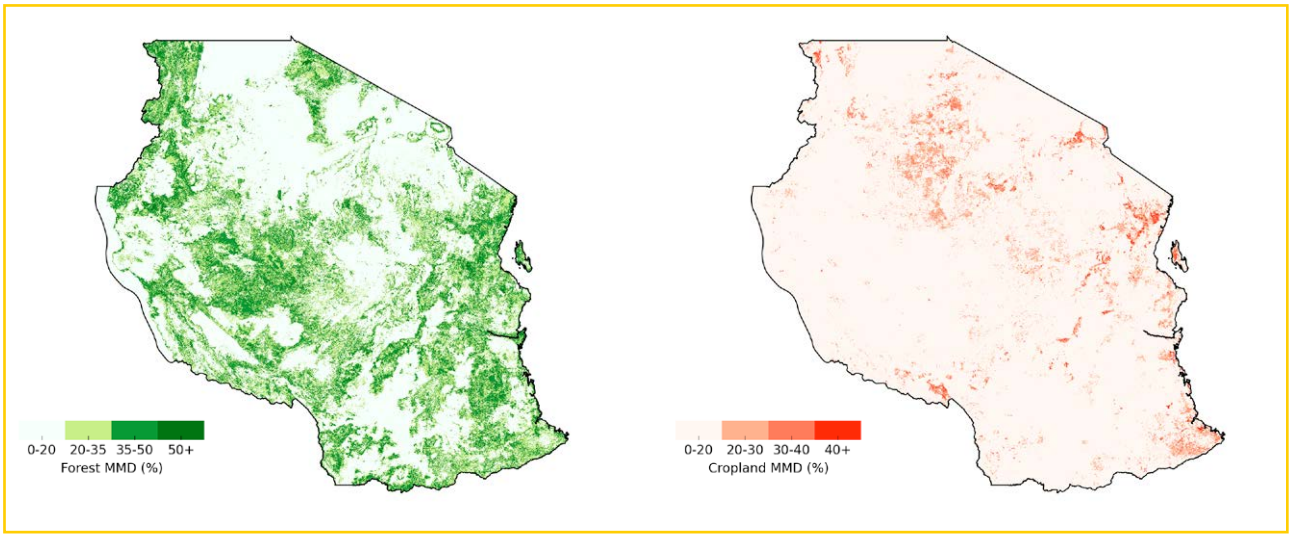


Figure 22. Average minimum measurable disagreement (MMD) in Oromia region between forest and cropland area estimates in data products.

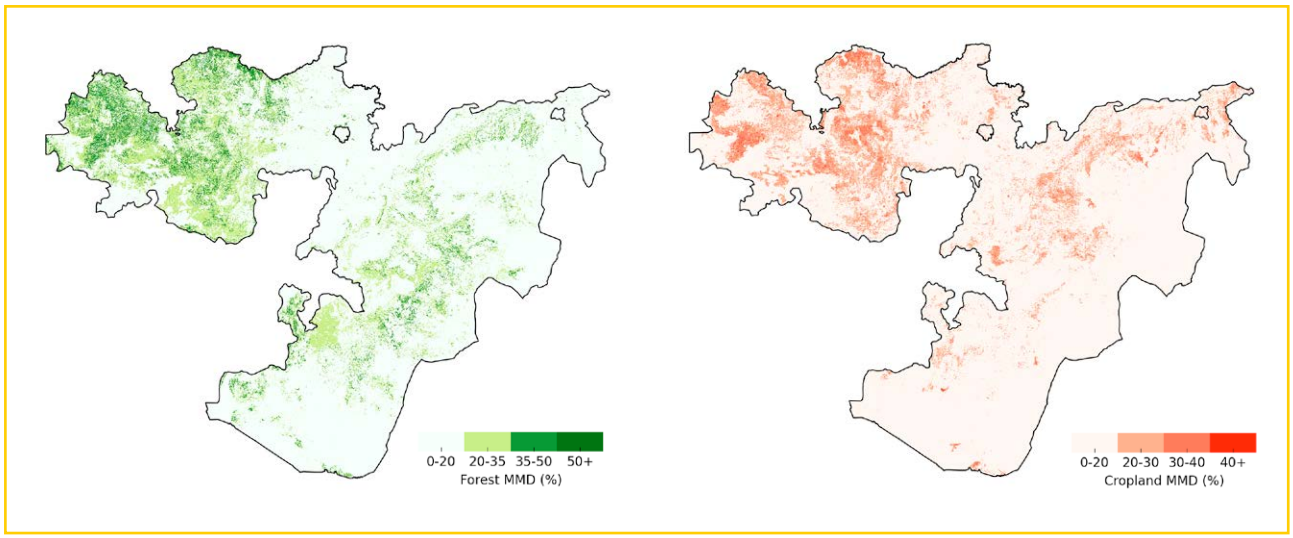


Figure 23. Average minimum measurable disagreement (MMD) in the western region of Ghana between forest and cropland area estimates in data products.

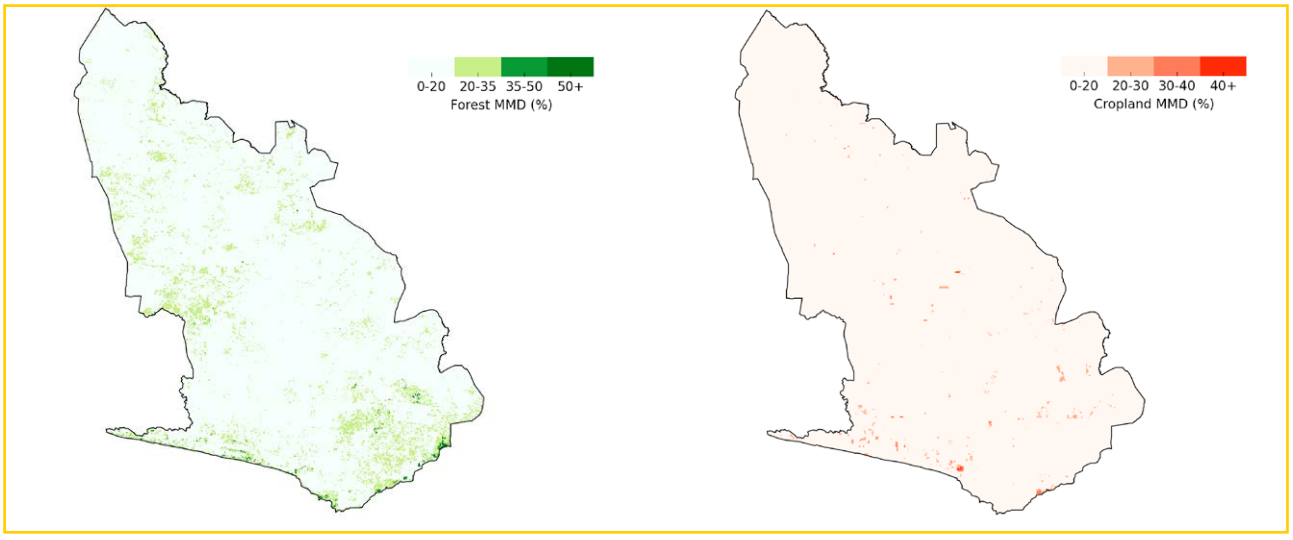
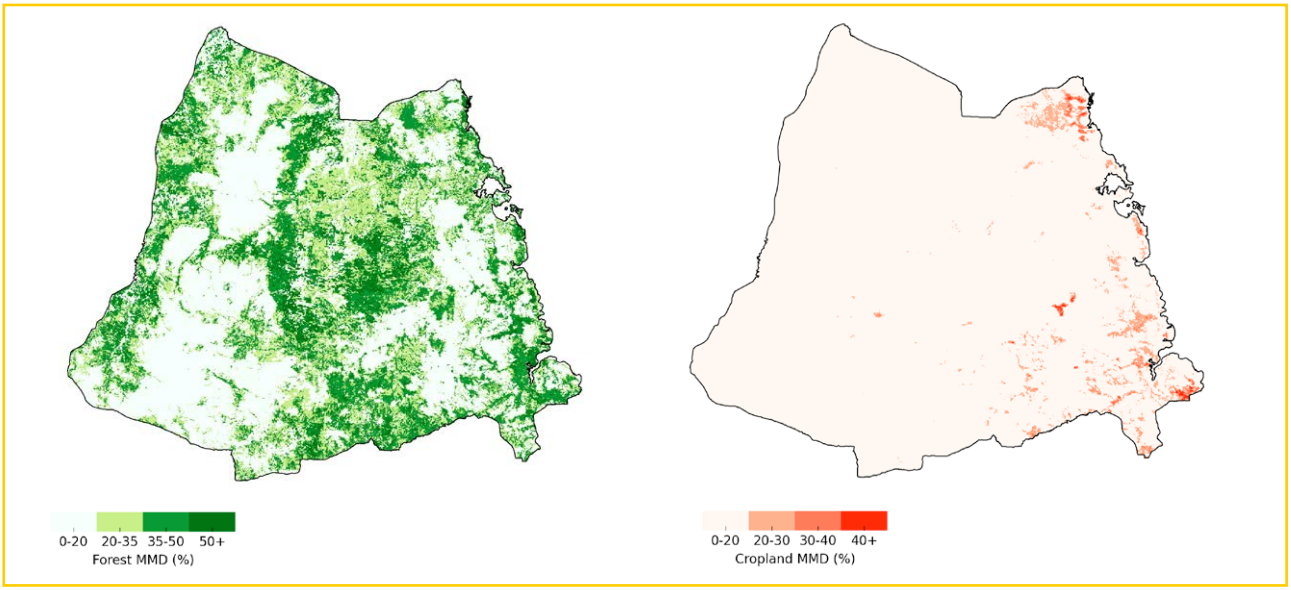


Figure 24. Average minimum measurable disagreement (MMD) in Lindi region between forest and cropland area estimates in data products.



3.3 Forest and cropland area change

There exist large differences in rates of forest and cropland area change between those data products that present data from multiple years (Table 4). For example, mean estimates of forest area change in sub-Saharan Africa vary from a loss of 0.50 per cent per year (FAOSTAT) to a gain of 0.88 per cent per year (JAXA). Meanwhile croplands are predicted to be expanding in sub-Saharan Africa at a rate averaging from 0.05 per cent per year (ESA CCI) to 1.87 per cent per year (GlobCover). Similar differences in predictions are also observed in national and regional data.

Graphing relative changes in forest and cropland area shows forest loss and cropland expansion across most study regions (Figures 25–31). This is consistent with widespread reports of forest loss and expansion of agriculture across sub-Saharan Africa. Some data products show fairly consistent gains/losses between years (eg FAOSTAT, Hansen), whilst others show large fluctuations between years which can be attributed to noise in data (eg MODIS VCF). In general, inter-year fluctuations in the forest and cropland area predictions of a single data product appear greater at regional scales than at the continental scale where noise is likely to cancel out across different areas.

Table 4. Mean annual rates of land use change (%/yr) for forest and cropland change estimated by data products with multiple measurements available through time.

	SUB-SAHARAN AFRICA		ETHIOPIA		GHANA		TANZANIA		OROMIA REGION		WESTERN REGION		LINDI REGION	
Data product	Forest	Crop-land	Forest	Crop-land	Forest	Crop-land	Forest	Crop-land	Forest	Crop-land	Forest	Crop-land	Forest	Crop-land
ESA CCI	−0.07	0.04	−0.40	0.03	−0.12	0.02	−0.06	0.04	−0.10	0.02	−0.07	0.03	−0.05	0.05
FAOSTAT	−0.50	1.22	−0.75	3.40	0.31	1.51	−0.79	3.58	–	–	–	–	–	–
GlobCover	0.69	1.87	4.15	1.06	−7.02	6.32	−0.07	0.34	4.67	−0.32	−4.69	8.45	−5.38	4.34
Hansen	−0.20	–	−0.08	–	−0.26	–	−0.26	–	−0.09	–	−0.65	–	−0.26	–
JAXA	0.88	–	−1.73	–	1.45	–	0.16	–	−1.58	–	−0.16	–	−2.88	–
MODIS LULC	−0.14	0.56	−0.25	2.04	−0.45	0.82	0.35	−0.96	0.76	2.34	0.01	0.91	0.60	−3.40
MODIS VCF	0.05	–	0.03	–	0.68	–	−1.82	–	−0.20	–	1.48	–	0.62	–
Mean average	0.10	0.92	0.14	1.63	−0.77	2.17	−0.36	0.75	0.58	0.68	−0.68	3.13	−1.22	0.33

Figure 25. Rates of land cover change measured in sub-Saharan Africa relative to forest/cropland area in the first measured year for each data product.

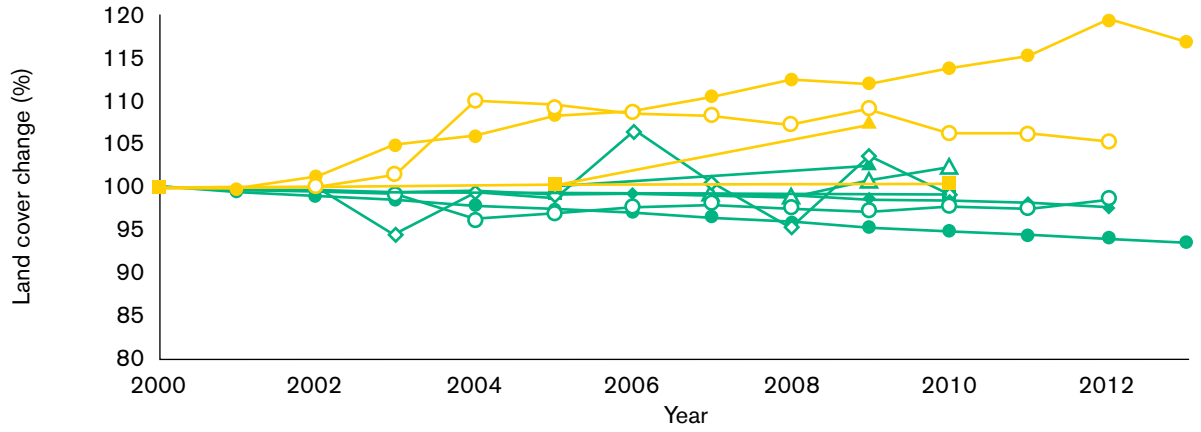


Figure 26. Rates of land cover change measured in Ethiopia relative to forest/cropland area in the first measured year for each data product.

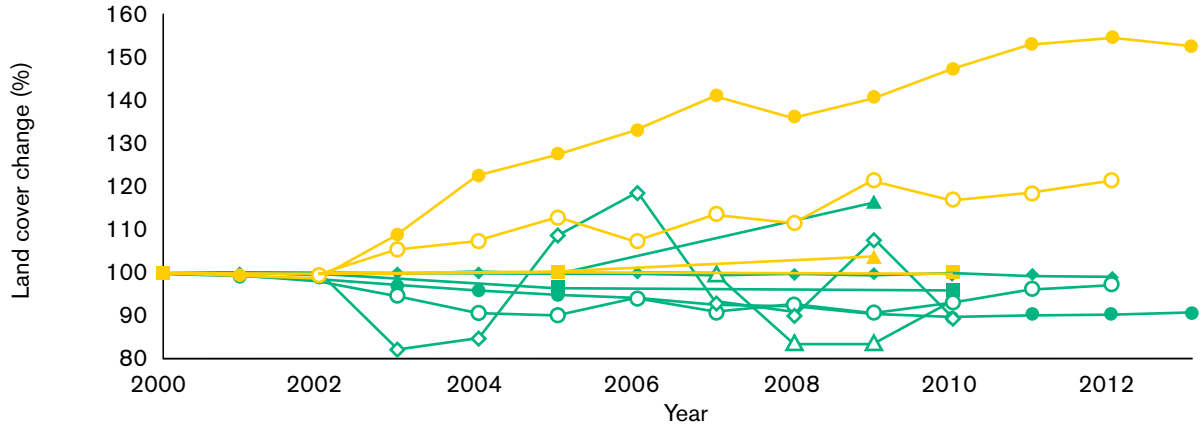
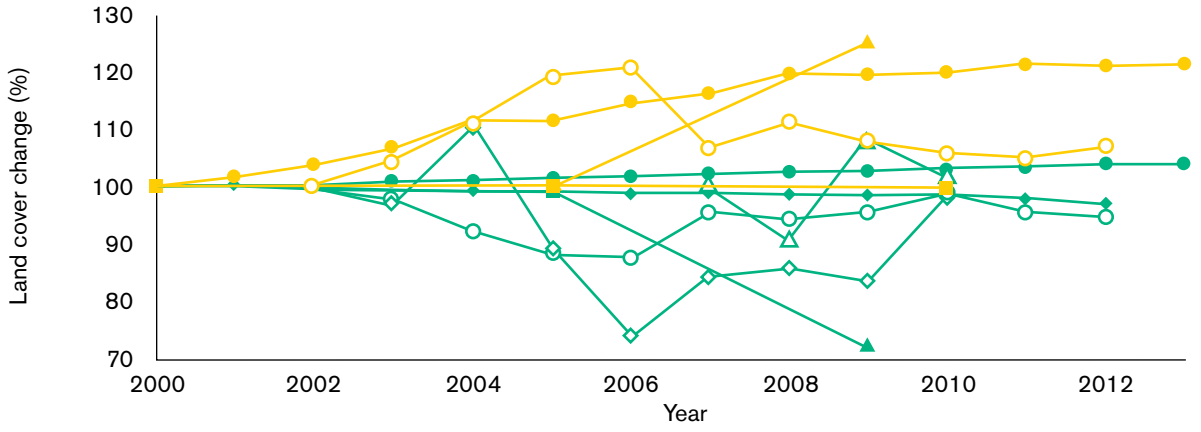


Figure 27. Rates of land cover change measured in Ghana relative to forest/cropland area in the first measured year for each data product.



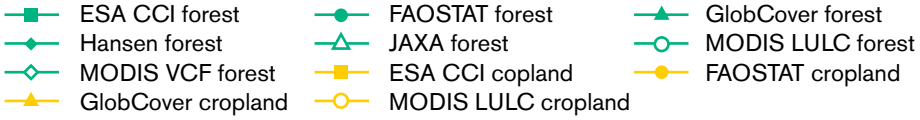


Figure 28. Rates of land cover change measured in Tanzania relative to forest/cropland area in the first measured year for each data product.

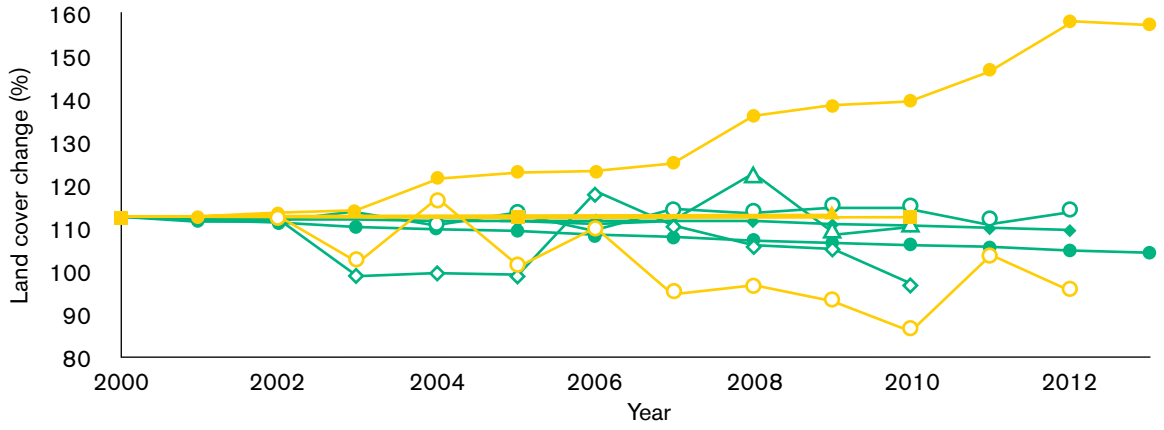


Figure 29. Rates of land cover change measured in Oromia region relative to forest/cropland area in the first measured year for each data product.

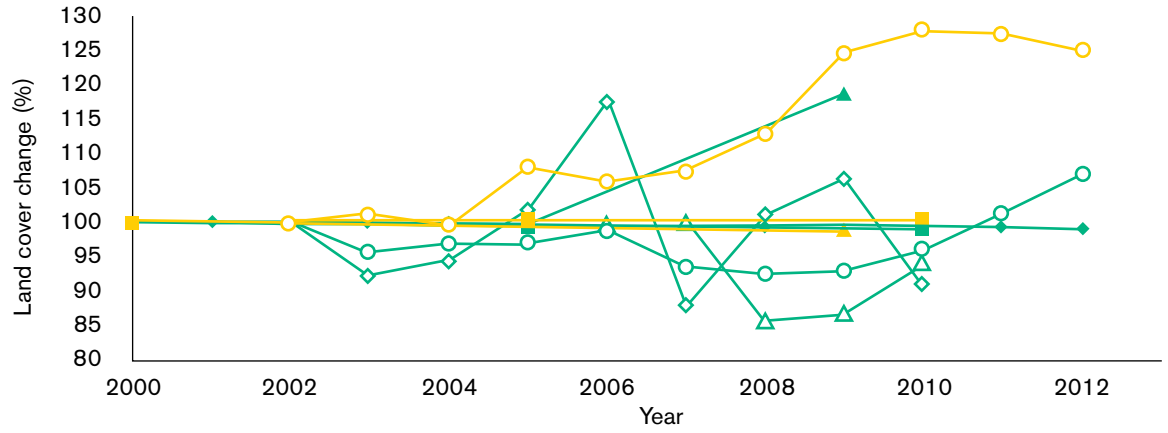


Figure 30. Rates of land cover change measured in the western region of Ghana relative to forest/cropland area in the first measured year for each data product.

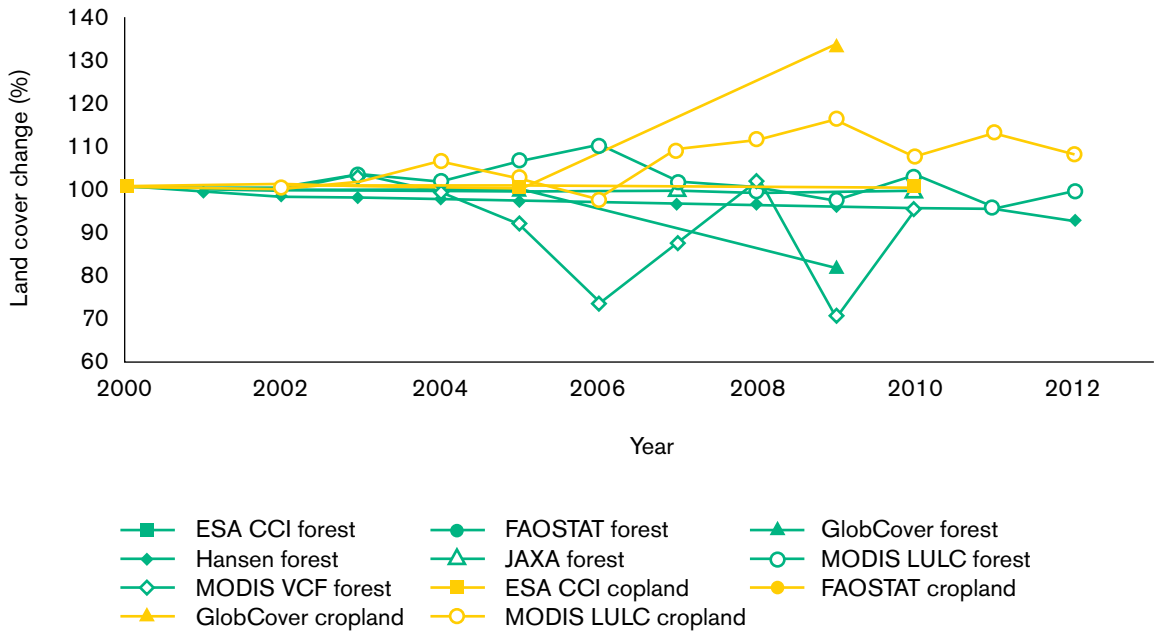
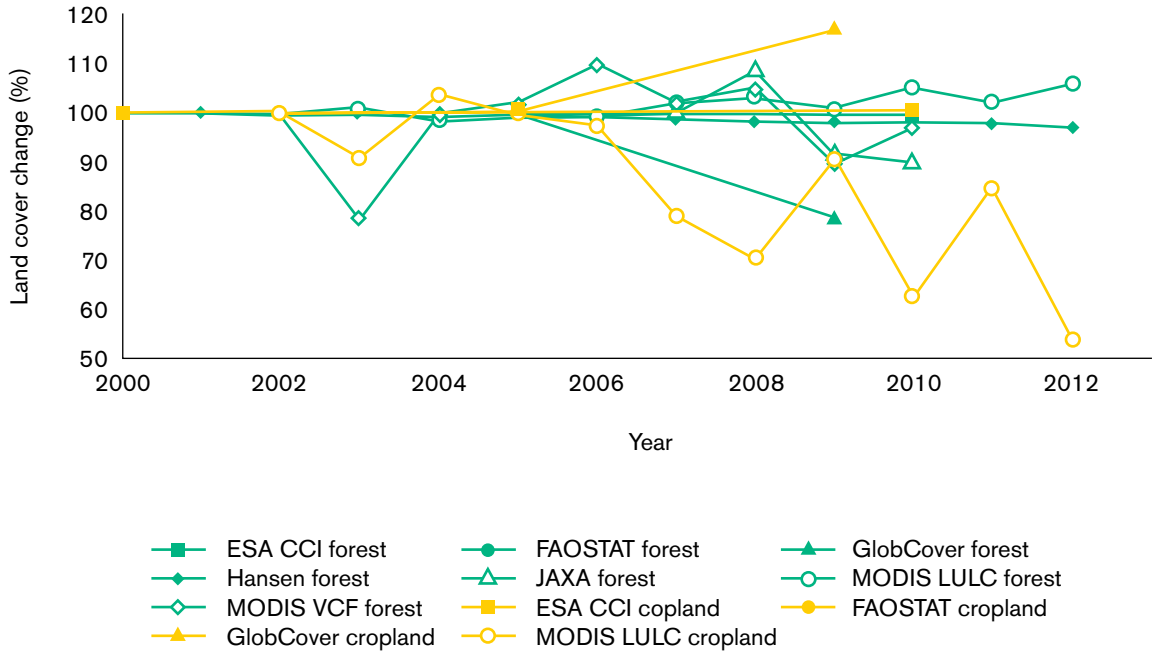


Figure 31. Rates of land cover change measured in Lindi region relative to forest/cropland area in the first measured year for each data product.



4

Discussion

Large areas of SSA are experiencing rapid deforestation in part the result of expansion of croplands for production of food and cash crops. Efforts to monitor the causes and consequences of deforestation are reliant upon the availability of accurate data describing the distribution of land cover types and rates of land cover change. In this report we compare the forest and cropland area estimates of 14 widely used data products to illustrate the high level of uncertainty associated with land cover data in Africa.

We found that data products to a large degree do not agree on the area, the spatial distribution, or the rates of change of forests and croplands in Africa. The area of forest and cropland cover estimated by data products varies widely, with the greatest disagreement apparent in estimates of cropland area, and where forest/cropland area is calculated at regional spatial scales. Disagreement between data products about the presence of forest land cover is notably elevated in savannah and woodland areas, with good agreement on the presence/absence of forest in areas of dense rainforest and non-vegetated land. Cropland products show good agreement in areas where croplands are rare or absent, though do not agree well in areas where croplands are present. Disagreement between data products is notably elevated where agriculture is small-scale and spatially spread such as in Tanzania. Where multiple years of land cover estimates are available land cover products largely agree that Africa is experiencing net losses of forest and net gains in cropland area. The rates of forest area loss and cropland area increases estimated by data products are not consistent with trajectories of land cover change and most uncertain at regional spatial scales.

Here we discuss some of the main sources of disagreement between data products, how existing data may be used to better understand the dynamics of forest loss and cropland expansion and the future prospects for improved land cover data in Africa.

4.1 Sources of disagreement between data products

Inconsistent and poorly defined classification schemes

There are a number of factors leading to inconsistency between data products. Perhaps the greatest of these is in harmonising inconsistent definitions of forests and croplands. As an example, there are a number of land cover classes in the GlobCover land cover product that may at least partially represent forest (Table 5). Where categories are well-defined as forest (ie classes 40, 50, 60, 70, 90, 100) a large proportion of the total area of the pixel can be attributed to forest. In the case of the GlobCover product these pixels represent 70–100 per cent forest cover. Other classes (ie 20, 30, 110, 120) are less clear; in these cases we must make assumptions about the proportion of the mixed classes that represent forest based on the class description, which can be very ambiguous. For instance, class 110 contains 50–70 per cent forest or shrubland, meaning that forest in that location may conceivably be 0–70 per cent of the present land cover. In all cases we assume the central value to represent the true forest cover (in this case 35 per cent) which implicitly assumes a uniform distribution of land cover in each mixed class. This is an assumption cannot be verified without external validation data.

Even where the forest/cropland area of a pixel can be unambiguously determined from class descriptors there remains a difficulty in resolving these classes into a consistent definition of forests or croplands. The absence of a universally recognised definition of forest or cropland results in a range of definitions being adopted in data products. This issue is well illustrated by a review of the various definitions of forest that found many hundreds of existing definitions are in operation around the world (Lund, 2008). Where land cover products use different definitions for what constitutes a forest or cropland area, these land cover types may not be fully comparable. This is further complicated by the frequently unclear definitions of land cover provided by documentation.

Resolving disparate class definitions into consistent land cover types is a well-known problem and efforts have been made to improve the comparability of data legends. Most prominently the land cover classification

Table 5. Land cover classes in GlobCover that represent forest and cropland.

VALUE	DESCRIPTION	PFOREST	PCROP
11	Post-flooding or irrigated croplands (or aquatic)	0	0.7 – 1
14	Rainfed croplands	0	0.7 – 1
20	Mosaic cropland (50-70%) / vegetation (grassland/shrubland/forest) (20-50%)	0 – 0.5	0.5 – 0.7
30	Mosaic vegetation (grassland/shrubland/forest) (50-70%) / cropland (20-50%)	0 – 0.7	0.2 – 0.5
40	Closed to open (>15%) broadleaved evergreen or semi-deciduous forest (>5m)	0.7 – 1	0
50	Closed (>40%) broadleaved deciduous forest (>5m)	0.7 – 1	0
60	Open (15-40%) broadleaved deciduous forest/woodland (>5m)	0.7 – 1	0
70	Closed (>40%) needleleaved evergreen forest (>5m)	0.7 – 1	0
90	Open (15-40%) needleleaved deciduous or evergreen forest (>5m)	0.7 – 1	0
100	Closed to open (>15%) mixed broadleaved and needleleaved forest (>5m)	0.7 – 1	0
110	Mosaic forest or shrubland (50-70%) / grassland (20-50%)	0 – 0.7	0
120	Mosaic grassland (50-70%) / forest or shrubland (20-50%)	0 – 0.5	0

Some classes might represent entirely forest ($pForest = 1$) and some mixed classes are only partially forest ($pForest < 1$), with the proportion of these pixels representing true forest cover remaining ambiguous given the descriptions of land cover. Tables of all land cover classes in each data product are reproduced in Appendix 1.

system (LCCS) applies a standardised set of classifiers and thresholds to aid the harmonisation of land cover datasets (FAO 2016). Where land cover products adopt LCCS standards (eg ESA CCI, GLC-2000, GLC-SHARE), their inter-comparison is made more straightforward.

In this work we aimed for consistency by adopting the commonly used FAO definitions of forest and cropland. With measurement of forest cover in particular, maintaining consistent definitions between land cover products required untested assumptions of the threshold tree cover (Hansen, MODIS VCF) or biomass values (Avitabile) that separate forests from non-forest. It is noteworthy that three of the four largest estimates of SSA forest extent come from such data products suggesting that this definition of forest is more inclusive than those in use by other data products. This is reflected in some of the greatest disagreement between forest area estimates being present in moderate biomass savannahs and woodlands, land cover types which are inconsistently defined as forests. We also note that 10 per cent tree cover and biomass of 10 tonnes of carbon per hectare does not preclude the presence of another land cover type (eg agriculture) which may result in these areas being given another classification on other data products.

Classification errors

Classification errors are a further important source of uncertainty, occurring when a pixel that is classified as one land cover type in reality belongs to another land cover type. Most data products report classification accuracies of 70–80 per cent (see Appendix 1), though it is acknowledged that accuracy varies spatially and errors may be higher or lower at a given location. Classification uncertainty is likely to be elevated on the African continent where data for training and validating land cover classification algorithms are extremely limited.

We observed that at several locations there was a high degree of agreement between different data products on the presence or absence of forests or croplands. These tend to be locations of tropical forest, high density agricultural land or bare ground (all land cover classes that fit unambiguously into given definitions of forest and cropland, and where the spectral properties of land cover types will be most distinct). Other areas show large disagreements between data products, most notably in savannah/woodland regions (eg Tanzania) and where agriculture is spatially dispersed amongst uncultivated land (eg Oromia region, Lindi region). As well as these areas being sensitive to differences in land cover definitions it is likely that they cause confusion to land cover classification algorithms by having spectral properties that are not easily discernible.

Classification error is also common between land cover classes that have similar spectral properties. An example of this is the lack of spectral distinction between croplands and grasslands or open canopy savannahs resulting in frequent confusion between locations with these two land covers. Uncertainty in classification often leads to spurious inter-annual variation in forest and cropland area predictions within a data product. For example, we observe large inter-annual differences in forest area predictions from the MODIS VCF and MODIS LULC data products which demonstrates how the land cover at a particular location can be differently classified even by the same classification algorithm.

Even where data products reach the classification accuracy of the best available land cover products (~80%), it is notable that the magnitude of this uncertainty is much greater than decadal rates of land use change observed to African forests and croplands (Figures 25–31). This means that detections of true land use change are likely frequently obscured by error in image classification. A number of data products mitigate this issue by limiting the freedom of pixels to change classification in cases where change is ambiguous (eg ESA CCI, MODIS LULC). In the case of the ESA CCI product this results in rates of land cover change that are in many locations unrealistically slow suggesting this data product has limited utility for the measurement of land cover change.

Issues of scale

Land cover data are available at a range of spatial scales; here we investigated data products that vary from 30 metres to 10 kilometres spatial resolution. Where the target land cover type tends to occur over areas that are smaller than the resolution of the data product systematic error in land cover area estimation may occur as a result of mixed pixels. Smallholder agriculture in southern Africa is common, with the result that croplands are very unlikely in many locations to be the dominant land cover type even in a 500 metre pixel. In these cases, the area of cropland that it represents may be missed. Where croplands exist as part of a vegetation mosaic uncertainty in classification may be particularly high, an effect we observe in cropland data in Tanzania. We expect that mixed pixels are less of a problem in high resolution forest area datasets (eg Hansen and JAXA) but high resolution cropland area datasets do not yet exist.

For this report we assessed the predictions of data products at continental, national and regional spatial scales. We found that disagreement between data products was more pronounced at regional spatial scales, with differences in estimates of forest and cropland area more likely to cancel out over continental spatial scales. This implies that land cover data suffer from localised biases with the result that estimates of forest and cropland area in a particular location may be associated with an uncertainty that is greater than that of the data product as a whole.

Further uncertainty in estimates of forest and cropland area is introduced by the temporal resolution of data products. Land cover data do not all represent the same time period, with our comparison of forest and cropland area using data from 2000 to 2012. This time period has seen a loss of forest and gain in cropland area in Africa so our assessment does not always compare like with like. Whilst the error introduced by this is likely to be small over large scales this can represent an important source of localised uncertainty where land cover change has been rapid.

There also exist challenges to the measurement of deforestation and cropland expansion from multi-temporal data that are particular to Africa. Much forest in SSA takes the form of savannahs or woodlands which have the propensity to switch dynamically from forest to grassland vegetation structures without having been deforested. The continuous gain and loss of forest area may result in overestimates of deforestation rates where regrowth is not taken into account. The dominance of shifting agriculture results in similar issues with cropland being created and abandoned continuously, and resulting cropland area and change estimates being prone to error.

4.2 The utility of land cover data for land use planning

Whilst uncertainty is significant, land cover datasets from remote sensing and census data offer one of the few methods to systematically map land cover over large scales and are an important source of data for land use planning. As a result of this review we make the following recommendations for use of current land cover data for monitoring deforestation and cropland expansion in Africa.

Carefully examine land cover definitions

It can be very difficult to harmonize different datasets to produce a coherent analysis of land cover. This is in large part the result of the divergent land cover definitions employed by different land cover data products. Thematic inconsistency leads to difficulty in determining exactly what a land cover product's predictions of deforestation and cropland expansion mean in reality. Where the land cover definitions employed by a data product are a poor match for the system of interest consider alternative sources of land cover data.

Select the most appropriate data products

The datasets we investigate here have a range of strengths and weaknesses. Radar datasets perform well in cloudy locations, optical datasets can identify areas of both forest and cropland and offer long time series, and census data have good thematic resolution. It is unlikely that there is a single best data product for measurement of forest loss and cropland expansion, with each data source potentially performing better in different locations where different classes are of interest

One way to determine which land cover product is most appropriate is to compare it with ground truth data at the study site. This need not necessarily involve visiting a location with many data products turning to expert interpretation of Google Earth imagery to generate training and validation data. Validation of a dataset in the intended study region gives the user a means to select the land cover map that best corresponds to the study region and the requirements of a project.

Think about scales of interest

Land cover datasets differ in their spatial, temporal and thematic resolution, and some may have properties that are unsuitable for a given application. For example, measurement of small-scale deforestation events or identification of croplands where fields are widely dispersed and small may be unfeasible without high spatial resolution data, and where agriculture is shifting reliable estimates of its extent and impacts will be limited where measurements of land cover are infrequent.

We found that uncertainty in predictions of forest and cropland area were particularly great when considered over small scales. In these cases, regional land cover data should be preferred over global land cover data as biases are likely to be locally significant in global land cover data.

Bear uncertainty in mind

There is a large uncertainty associated with estimates of forest and cropland area in available global land cover products. This means that locations identified as forest or cropland are not necessarily forest or cropland, and locations that are identified as neither forest nor cropland may in reality be forest or cropland. Similarly, locations where a data product reports a loss of forest does not necessarily indicate deforestation, and where a data product reports the appearance of cropland this does not necessarily indicate cropland expansion. These issues are particularly true when working with data at regional or local spatial scales where agreement between data products is poor. Where uncertainties and biases in land cover estimates are large, an uncritical analysis of forest loss and cropland expansion may result in misleading conclusions.

Consider a bespoke land cover product

The same techniques that are used to generate global land cover datasets can be applied at smaller spatial scales to generate regional land cover maps. Whilst this is a considerable investment of effort the results are likely to be improved by being optimised to the location and land covers of interest. Furthermore, such data can be validated using local observations with the result that uncertainties will be much better understood than if using a global land cover product.

There may be limitations to achievable thematic resolution

For many important questions relating to land use change the information available from remote sensing data products may be insufficient. One example is identification of the crop types that are the main causes of forest loss, data which is not widely available in Africa where even estimates of the spatial extent of croplands are uncertain. Where remote sensing data are not able to distinguish between crop types a more useful approach may be to combine census data with available remote sensing observations (eg MapSPAM).

4.3 Future developments in land cover data

The generation of accurate and globally consistent land cover maps is recognised to be of great importance, and considerable institutional effort, computational processing power and investment in new technologies are being committed to improving land cover data. The result is that the accuracy and reliability of forest and cropland area mapping will improve over the coming years.

The mapping of global forest extent has developed rapidly over the past decade with the release of global biomass maps (eg Avitabile) and high resolution deforestation estimates (eg Hansen) - both important advances. Over the coming decade the mapping of forest land cover will benefit from a number of new satellite technologies. The European Space Agency's (ESA) Sentinel missions are a series of operational satellites that provide a range of optical and radar data. This new data will be of relatively high resolution (10m) and with open access will be a valuable future source of data for environmental monitoring. The global ecosystem dynamics investigation lidar (GEDI), to be launched in 2018, will measure the vertical structure of global temperate and tropical forests (Dubayah *et al.* 2014). By measuring canopy height to an accuracy of about one metre, GEDI will be used to generate global estimates of above ground biomass that will improve estimates of global above ground biomass (eg Avitabile). The ESA BIOMASS mission, to be launched 2020, is a radar satellite that will measure above ground biomass of the world's forests (Le Toan *et al.* 2011). It will improve upon similar measurements by the advanced land observing satellite (ALOS) and the recently launched ALOS-2 by using a longer wavelength which is more sensitive to differences in biomass in dense tropical forests. Data from these instruments will provide new and more detailed information for estimation of forest area and the consistent and high frequency mapping of deforestation events.

The development of technologies for the mapping of croplands is, in many ways, behind that for forests. This is reflected in the current availability and quality of data for cropland area estimation which are more limited and more likely to disagree than forest cover data. Whilst the future mapping of cropland extent will benefit from new

data from satellite sensors (eg ESA Sentinels), a major limitation at present is in the availability of appropriate data to train and validate the predictions of land cover classification algorithms. With computational algorithms still yielding unreliable results, efforts to improve cropland area datasets have focused on crowdsourcing of data collection to map individual fields (eg http://mappingafrica.princeton.edu/ and www.geo-wiki.org/). Such data can also be used to identify which dataset best represent crop area for hybrid data products and may be used as training data to improve cropland area detection algorithms.

Further to the problem of mapping cropland area there is interest in identification of the individual crop types that constitute agricultural land. Where it is already difficult to distinguish the spectral properties of croplands alone, separation of croplands into categories is a challenging problem. One approach is the spatial apportioning of agricultural production statistics to existing estimates of crop suitability and cropland land cover maps. This is the method used by MapSPAM (You *et al.* 2014), a data product that estimates the distribution of 42 crop types across the world based on the IIASA-IFPRI cropland cover map. The use of remote sensing methods to delineate crop types may in future be applicable to Africa, though these are limited to locations where the land area covered by cropland is better understood. For example, there now exists a map of eight UK crop varieties with a reported accuracy of 96 per cent based on remote sensing data from the ESA Sentinel satellites (www.ceh.ac.uk/crops2015). This is achievable by using detailed field boundary data from Ordinance Survey, consistent cartographic data that are far in advance of that available on the African continent.

5

Concluding remarks

Global land cover data are a useful, though uncertain, gauge of land cover patterns on the African continent. In this work we identified uncertainty in estimates of forest and cropland area in SSA across multiple spatial scales by comparing the land cover estimates of a range of land cover products. Large uncertainties are attributable to poor classification accuracy, imprecision in land cover definitions, and limitations to the spatial and temporal resolution of current land cover data.

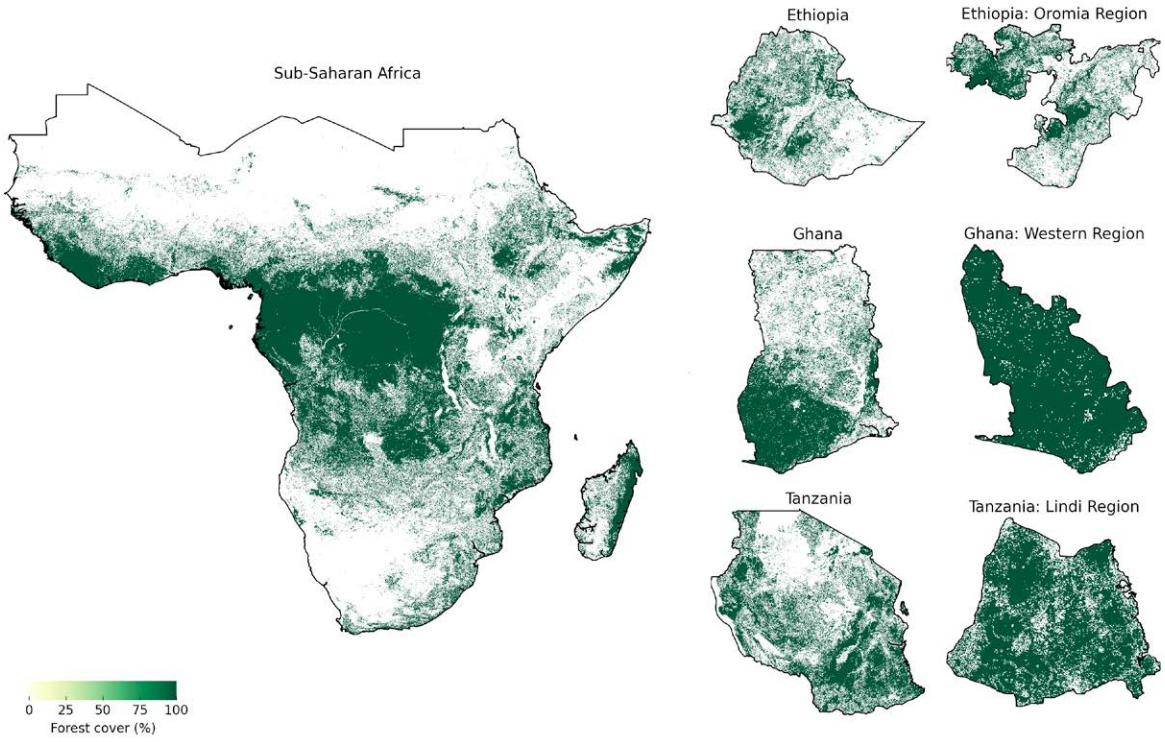
As a result of the limitations discussed in this paper, it is difficult to reliably measure the linked processes of cropland expansion and deforestation or to answer many important questions relating to the agriculture drivers of deforestation. For example, for companies to monitor the impacts of specific crops on forests in their supply chain, or for governments to manage the trade-offs between meeting domestic food demand and forest conservation, more targeted and accurate monitoring of the impacts of different crop types on forests is required.

However, there are ways to overcome some of the limitations. Data users can carefully examine land cover definitions, and choose those data products that are best suited to serve their objectives. Investment in generating custom land cover datasets at national or regional spatial scales, using a combination of remote sensing technology and the validation of local observations, can also result in improved land cover estimates.

Over the coming decade advances in satellite technologies will continue to improve the data available with higher resolutions and greater sensitivities to differences in biomass. Scarcity of data on the ground to calibrate and test remote sensing data products is being tackled through crowdsourcing of data to capture land use and land use changes that occur over small areas. Until then, our land cover data sources only offer limited answers and it is important to use the spatial information they generate with caution.

Appendix 1: Documentation for assessed data products

Avitabile



Description

The carbon stored in forest biomass is a key component of the global carbon cycle, though the size of this carbon pool and its spatial variability is unclear contributing a large uncertainty to estimates of carbon emissions due to land use change (Houghton *et al.* 2012). Uncertainty is particularly great in tropical regions, where observational data are rare and ecological understanding is limited. Efforts to reduce this uncertainty have integrated various satellite observations to generate pan-tropical estimates of forest biomass (Saatchi *et al.* 2011, Baccini *et al.* 2012). These maps were both generated moderate resolution imaging spectrometer (MODIS) data, using algorithms trained using light detection and ranging (LiDAR) data from the geoscience laser altimeter system with a series of forest inventory plots. The Avitabile biomass map is an integrated version of the Saatchi and Baccini biomass maps (Avitabile *et al.* 2016).

Thematic, spatial and temporal resolution

The Avitabile biomass map represents above ground biomass (AGB), which varies from 0 to 600 tonnes of carbon per hectare (tC/ha). The spatial resolution of biomass estimates is 1 kilometre. A single estimate is available which utilises data spanning the years 2000 to 2013.

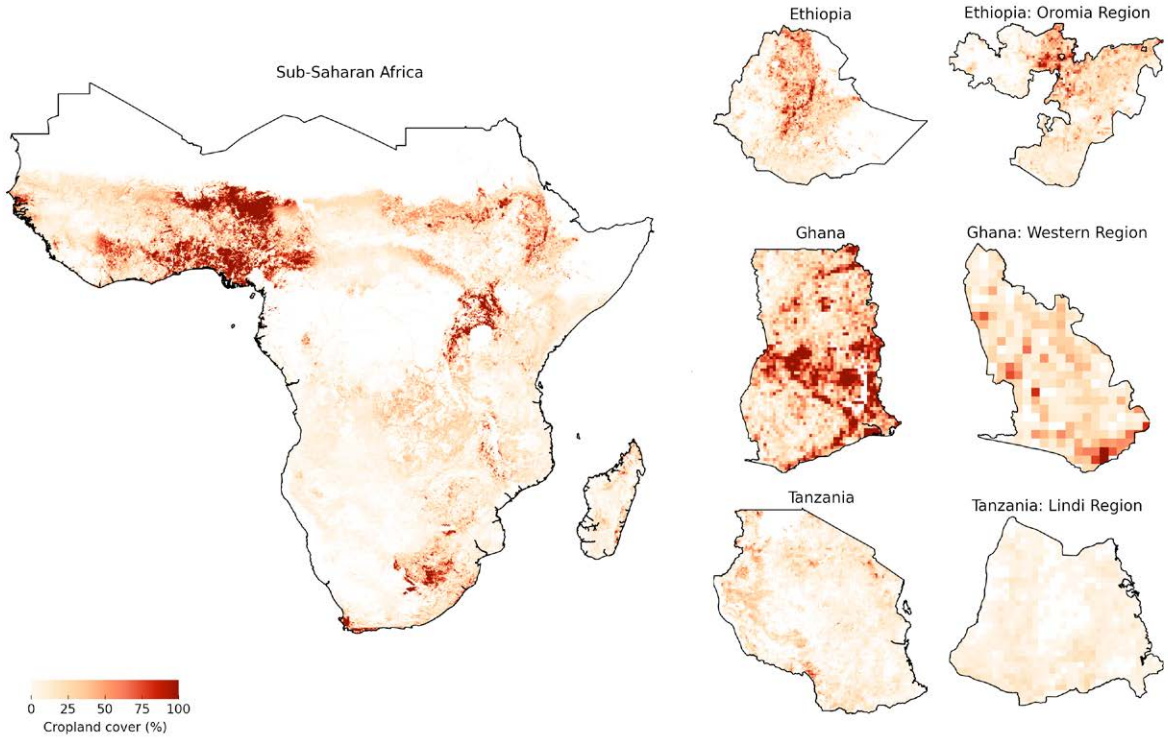
Representation of forest and cropland

AGB estimates can be used to delineate forested from non-forested areas by selection of a threshold biomass above which we assume forested vegetation. Using measurements from southern African woodlands, we set this threshold is set at 10 tC/ha, above which woodlands tend to meet to FAO definition of forests having greater than 10% canopy cover. Use of a single threshold value delineating forests can only be considered to result in a rough estimate of forest area as this value is poorly quantified and can be expected to differ between different ecosystems.

Accuracy assessment

The original pan-tropical biomass maps of Saatchi and Baccini show notable differences in AGB patterns at regional scales (Mitchard *et al.* 2013). Validation of the fused 'Avitabile' map shows a reduced error relative to the two inputs maps and importantly a considerably lower bias. However, uncertainty in this data product remains large with limited training and validation data and use of an inflexible threshold biomass for forest delineation resulting in particular uncertainty.

Croplands 2000



Description

The dataset presented in Ramankutty *et al.* (2008) Croplands 2000 estimates the global spatial distribution of agricultural lands by merging inventory data with satellite estimates of cropland area. Census data describing cropland were compiled from 57 countries at national and subnational scales. African countries are poorly represented in this database. Estimates from census data are made spatially-explicit using satellite estimates of cropland locations are taken from the MODIS LULC land cover product (Friedl *et al.* 2002) and the GLC-2000 land cover map (Bartholomé and Belward 2005), with the census data presumed to represent the 'true' area of agricultural land. The method for merging this data is described in Ramankutty and Foley (1998) and Ramankutty *et al.* (2008).

Thematic, spatial and temporal resolution

The data cover the globe with a spatial resolution of 5 arcminutes (~10km) and represent cropland area in the year 2000.

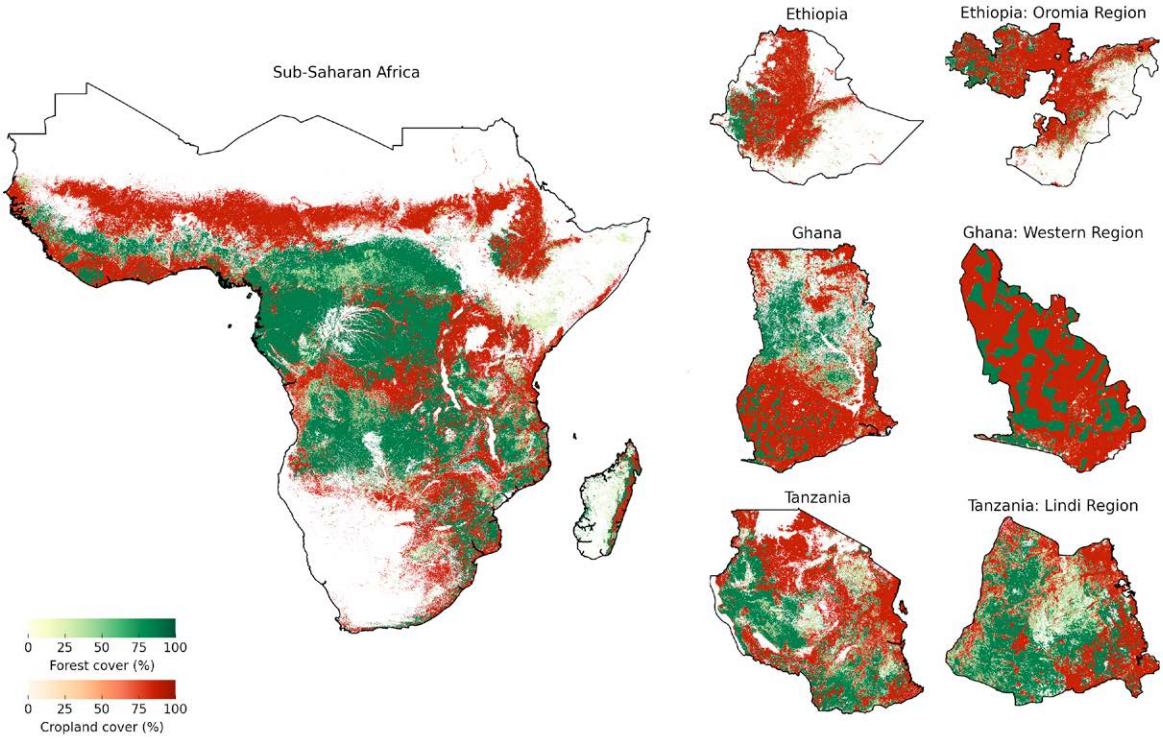
Representation of forest and cropland

Croplands were defined with the FAO definition of 'arable lands and permanent crops'. Data are reported as the proportion of each pixel that is covered by cropland. Forest area is not estimated by this data product.

Accuracy assessment

The authors report there are 15 million km^2 of cropland globally (12% of the Earth's ice-free surface), with a 90% confidence range of 12.2–17.1 million km^2 of cropland area. Qualitative comparisons with other data products were performed. Of particular relevance is the comparison with Africover data (Latham 2001) in East Africa, where cropland area was observed to compare 'reasonably well'.

ESA CCI



Description

The ESAs climate change initiative (CCI) land cover product is a recently released global land cover map generated using optical data from the medium resolution imaging spectrometer (MERIS) satellite sensor. The ESA CCI data product is designed for use in climate modelling studies, with an intention to extend the record back to the 1980s with data from the Advanced Very High Resolution Radiometer (AVHRR) and extend forward with Sentinel-3 and Proba-V data. The ESA CCI data product is based on classification of surface reflectance data using methods that were developed from the GlobCover processing chain (Bontemps *et al.* 2013). The Earth's surface is stratified into 'equal reasoning areas' where ecology and remote sensing data are expected to be similar allowing for the regional tuning of classification parameters. Land cover maps for different years are not generated independently, instead being derived relative to a 'baseline' land cover map to identify land cover trajectories, a process which limits the impact of noise in the detection of deforestation events.

Thematic, spatial and temporal resolution

The ESA CCI data product provides three land cover estimates for the periods 1998 to 2002, 2003 to 2007 and 2008 to 2012 at 300 metre spatial resolution. The ESA CCI data product has 36 land cover classes, which are organised hierarchically into 22 'level 1' classes (eg 10, 60), and 14 'level 2' classes (eg 11, 61) which are used where regional information is of sufficient quality. Classes are designed to be compatible with the United Nations LCCS. Classes in the ESI CCI data product are:

VALUE	DESCRIPTION	PFOREST	PCROP
0	No Data	0	0
10	Cropland, rainfed	0	0.7 – 1
11	Cropland, rainfed: Herbaceous cover	0	0.7 – 1
12	Cropland, rainfed: Tree or shub cover	0	0.7 – 1
20	Cropland, irrigated or post-flooding	0	0.7 – 1
30	Mosaic cropland (>50%) / natural vegetation (tree, shrub, herbaceous cover) (<50%)	0 – 0.5	0.5 – 1
40	Mosaic natural vegetation (tree, shrub, herbaceous cover) (>50%) / cropland (<50%)	0 – 0.7	0 – 0.5
50	Tree cover, broadleaved, evergreen, closed to open (>15%)	0.7 – 1	0
60	Tree cover, broadleaved, deciduous, closed to open (>15%)	0.7 – 1	0
61	Tree cover, broadleaved, deciduous, closed to open (>15%): Tree cover, broadleaved, deciduous, closed (>40%)	0.7 – 1	0
62	Tree cover, broadleaved, deciduous, closed to open (>15%): Tree cover, broadleaved, deciduous, open (15–40%)	0.7 – 1	0
70	Tree cover, needleleaved, evergreen, closed to open (>15%)	0.7 – 1	0
71	Tree cover, needleleaved, evergreen, closed to open (>15%): Tree cover, needleleaved, evergreen, closed (>40%)	0.7 – 1	0
72	Tree cover, needleleaved, evergreen, closed to open (>15%): Tree cover, needleleaved, evergreen, open (15–40%)	0.7 – 1	0
80	Tree cover, needleleaved, deciduous, closed to open (>15%)	0.7 – 1	0
81	Tree cover, needleleaved, deciduous, closed to open (>15%): Tree cover, needleleaved, deciduous, closed (>40%)	0.7 – 1	0
82	Tree cover, needleleaved, deciduous, closed to open (>15%): Tree cover, needleleaved, deciduous, open (15–40%)	0.7 – 1	0
90	Tree cover, mixed leaf type (broadleaved and needleleaved)	0.7 – 1	0
100	Mosaic tree and shrub (>50%) / herbaceous cover (<50%)	0 – 0.7	0
110	Mosaic herbaceous cover (>50%) / tree and shrub (<50%)	0 – 0.5	0
120	Shrubland	0	0
121	Shrubland: Evergreen shrubland	0	0
122	Shrubland: Deciduous shrubland	0	0
130	Grassland	0	0
140	Lichens and mosses	0	0
150	Sparse vegetation (tree, shrub, herbaceous cover) (<15%)	0	0
152	Sparse vegetation (tree, shrub, herbaceous cover) (<15%): Sparse shrub (<15%)	0	0
153	Sparse vegetation (tree, shrub, herbaceous cover) (<15%): Sparse herbaceous cover (<15%)	0	0
160	Tree cover, flooded, fresh or brakish water	0	0
170	Tree cover, flooded, saline water	0	0
180	Shrub or herbaceous cover, flooded, fresh/saline/brakish water	0	0
190	Urban areas	0	0
200	Bare areas	0	0
201	Bare areas: Consolidated bare areas	0	0
202	Bare areas: Unconsolidated bare areas	0	0
210	Water bodies	0	0
220	Permanent snow and ice	0	0

Representation of forest and cropland

Forests in the ESA CCI product are represented by classes 50, 60–62, 70–72, 80–82, and 90. In addition four mosaic classes exist, which are classes 30, 40, 100 and 110. Forest classes in the tropics are the only pixels where land cover is updated between years. Croplands are represented by classes 10–12 and 20, as well as a fraction of mosaic classes 30 and 40.

Accuracy assessment

Accuracy of the ESA CCI dataset was assessed with expert interpretation of reference data from high spatial resolution imagery and NDVI profiles from the SPOT VEGETATION time series. An accuracy of 70.8–74.4% is reported, allowing for differences in classification where classes may be in agreement.

FAOSTAT

Description

The FAO statistics division (FAOSTAT, see http://faostat.fao.org/) report a range of data on global land cover and land use. The aim is to support standardised methods across the globe to monitor a range of largely economic statistical indicators are available describing agriculture, forestry and fisheries, land and water resources and use, climate and environment.

Thematic, spatial and temporal resolution

FAOSTAT data are available on national-scales over annual time-scales. Data are not always collected using the same methods and may use differing definitions of land cover. Data are not spatially explicit.

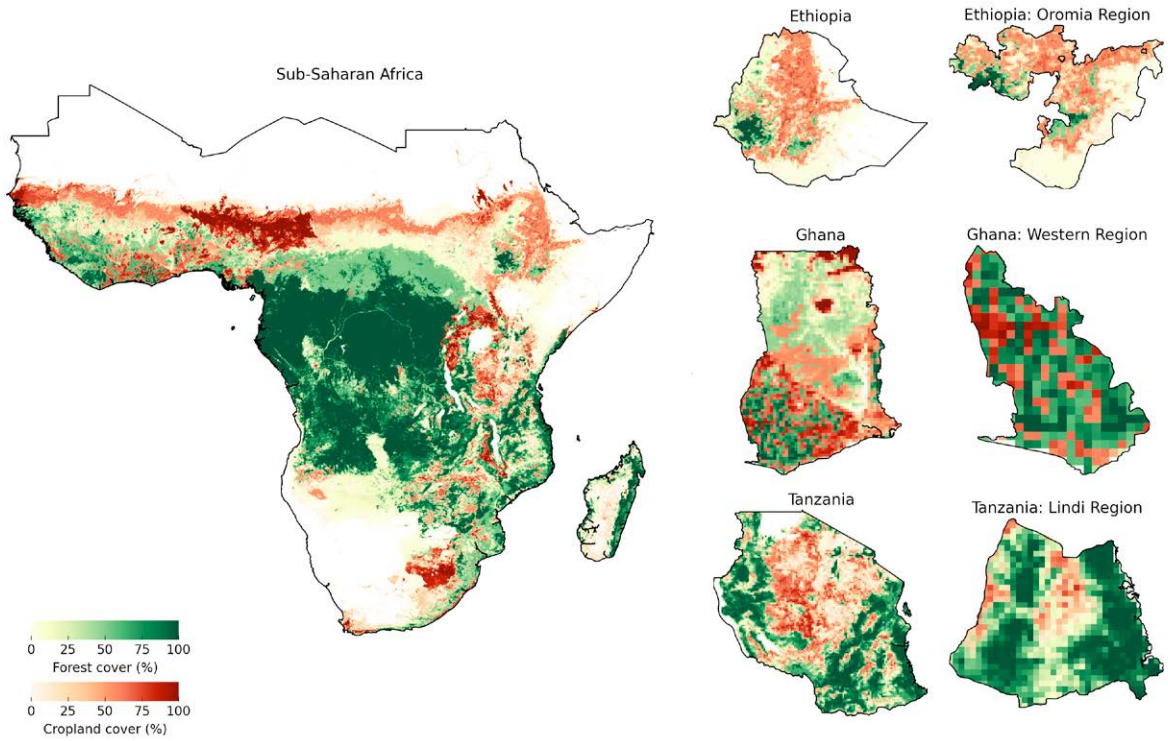
Representation of forest and cropland

FAOSTAT produces statistics for forest area and cropland area. These aim to follow the FAO definition of forest and cropland. For cropland area estimates we include arable land and permanent crops and exclude permanent meadows and pastures.

Accuracy assessment

The reliability of FAO data varies between countries and there are well-known discrepancies between this data and remote sensing measurements. FAO data are provided by national agencies which often lack the technical capacity to accurately assess forest and cropland area. Furthermore, definitions of these land cover types differ between countries resulting in uncertainties that are regionally variable.

GAEZ



Description

The Global Agro-Ecological Zones (GAEZ) system has been developed by the FAO in collaboration with the International Institute for Applied Systems Analysis (IIASA). The aim of GAEZ is to characterise the climate, soil and terrain conditions relevant to crop production, and identify potential and sustainable crop yields. Land cover data in GAEZ is derived from various remotely-sensed datasets: GLC-2000, the IFPRI agricultural extent database, the Global Forest Resources Assessments (FRA2000, FRA2005), the Digital Global Map of Irrigated Areas and a population data inventory for the year 2000.

Thematic, spatial and temporal resolution

GAEZ has a nominal resolution of 10km and provides a single estimate of land cover representing the year 2000. GAEZ includes maps of a wide range of other parameters relating to land resources, agro-climatic resources, sustainable and potential yields, actual yields and production, and yield and production gaps.

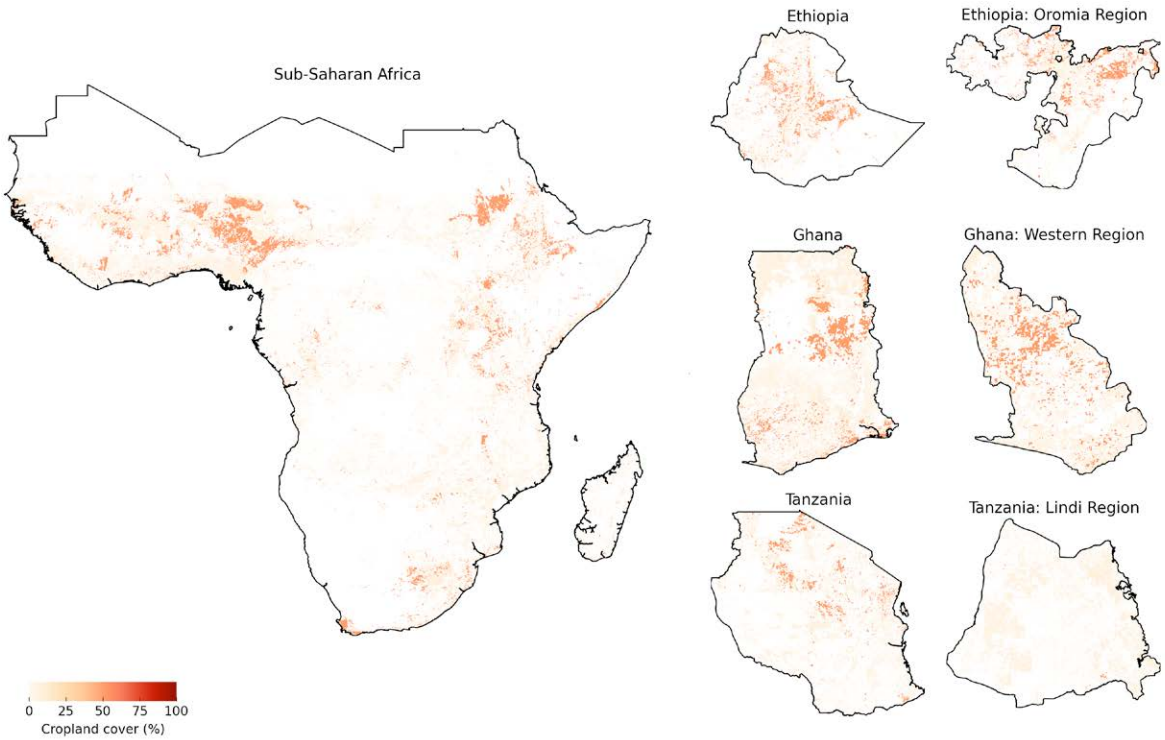
Representation of forest and cropland

Cropland and forest area are presented as the proportion of each 10km grid cell that is made up by forest or cropland cover.

Accuracy assessment

GAEZ does not provide an independent quantitative accuracy assessment, with accuracy best described by that of its constituent data products.

GCAD



Description

The Global Cropland Area Database (GCAD) is a remote-sensing based cropland area product under development by the United States Geological Survey (USGS). Whilst the eventual aim is to generate a global 30m resolution cropland area map, there are a number of products that have been generated at moderate spatial resolution and with varied spatial-extents (https://croplands.org/). Here we use the interim 1km resolution global cropland area product described in Teluguntla *et al.* (2015). This data product is generated though the synthesis of data from four studies: Thenkabail *et al.* (2009), Friedl *et al.* (2010), Pittman *et al.* (2010), Yu *et al.* (2013).

Thematic, spatial and temporal resolution

GCAD provides a single estimate of cropland area is available representing crop area from around the year 2000. The GCAD product describes cropland area, split into irrigated and rainfed agriculture, with information on dominance of cropland in each pixel. In addition to an improved spatial resolution, future data also aims to describe the relative dominance of the eight most common global crop types.

Representation of forest and cropland

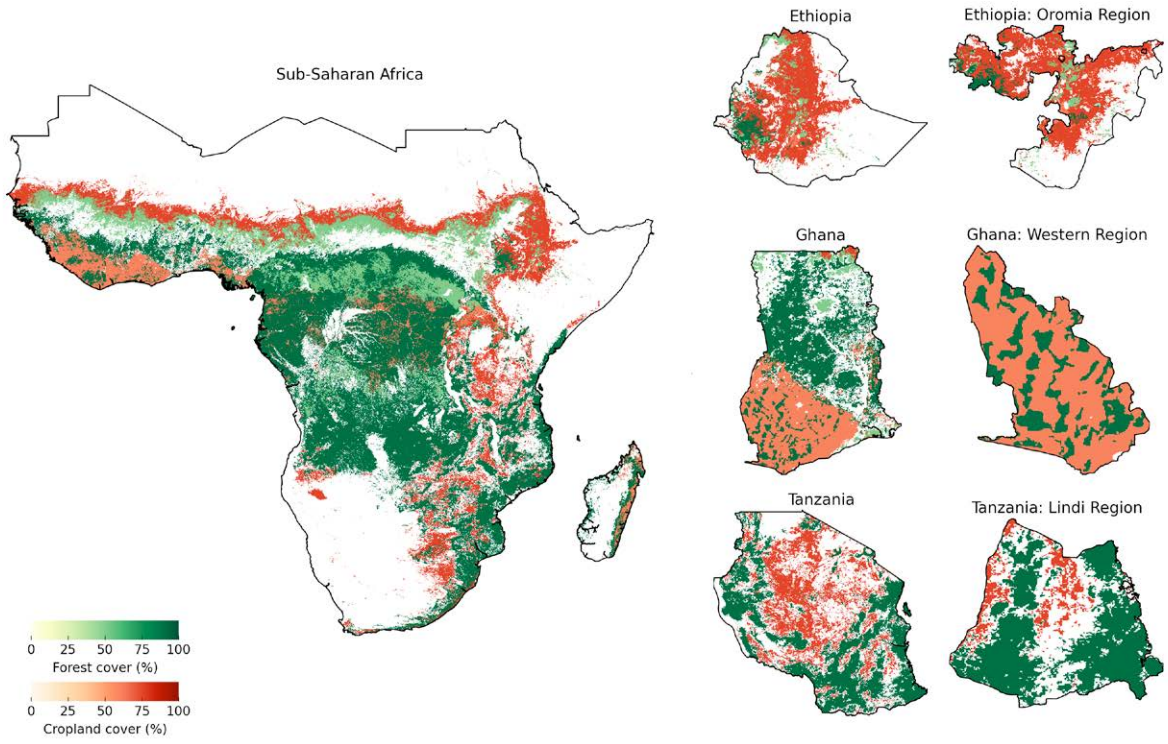
Croplands are defined as a minimum of 50% coverage in each pixel, which are split into irrigated and rainfed classes. Rainfed croplands have additional classes of minor and very minor fragments, which we assume to represent 10–20% and 0–10% cropland cover respectively. The GCAD dataset does not estimate forest area.

VALUE	DESCRIPTION	PFOREST	PCROPLAND
0	Not cropland	0	0
1	Croplands, irrigation major	0	0.5 – 1
2	Croplands, irrigation minor	0	0.5 – 1
3	Croplands, rainfed	0	0.5 – 1
4	Croplands, rainfed minor fragments	0	0.1 – 0.2
5	Croplands, rainfed very minor fragments	0	0 – 0.1

Accuracy assessment

The authors' note that the data from the four studies they synthesised to a large extent do not agree on cropland area, suggesting a large uncertainty associated with using a single product. As a product still under development no quantitative analysis of uncertainty has yet been performed.

GLC-2000



Description

The GLC2000 project aimed to produce a standardised global land cover product for the year 2000 explicitly linked to existing regional datasets (Bartholomé and Belward 2005). GLC2000 used the expertise of over 30 research teams, each of which provided a land cover map for a particular geographic region of interest. We use the land cover map developed for the African continent (Mayaux *et al.* 2004), which is based largely on optical data from the SPOT VEGETATION satellite. Monthly data were separated into 100 classes using an unsupervised clustering algorithm which were aggregated and labelled using visual interpretation, thematic maps and spectral statistics.

Thematic, spatial and temporal resolution

GLC-2000 data record vegetation properties in the year 2000 with a spatial resolution of 32 arcseconds (~1km). The GLC-2000 data product has 22 land cover classes derived from the FAO LCCS, which are selected to be inter-regionally comparable. The land cover classes in the Africa land cover dataset are:

VALUE	DESCRIPTION	PFOREST	PCROP
0	Background	0	0
1	Closed evergreen lowland forest	0.8 – 1	0
2	Degraded evergreen lowland forest	0.8 – 1	0
3	Submontane forest (900–1500m)	0.8 – 1	0
4	Montane forest (>1500m)	0.8 – 1	0
5	Swamp forest	0	0
6	Mangrove	0	0
7	Mosaic Forest / Croplands	0.2 – 0.8	0.2 – 0.5
8	Mosaic Forest / Savanna	0.2 – 0.8	0
9	Closed deciduous forest	0.8 – 1	0
10	Deciduous woodland	0.8 – 1	0
11	Deciduous shrubland with sparse trees	0	0
12	Open deciduous shrubland	0	0
13	Closed grassland	0	0
14	Open grassland with sparse shrubs	0	0
15	Open grassland	0	0
16	Sparse grassland	0	0
17	Swamp bushland and grassland	0	0
18	Croplands (>50%)	0	0.5 – 1
19	Croplands with open woody vegetation	0.2 – 0.8	0.2 – 0.8
20	Irrigated croplands	0	0.5 – 1
21	Tree crops	0	0.5 – 1
22	Sandy desert and dunes	0	0
23	Stony desert	0	0
24	Bare rock	0	0
25	Salt hardpans	0	0
26	Waterbodies	0	0
27	Cities	0	0

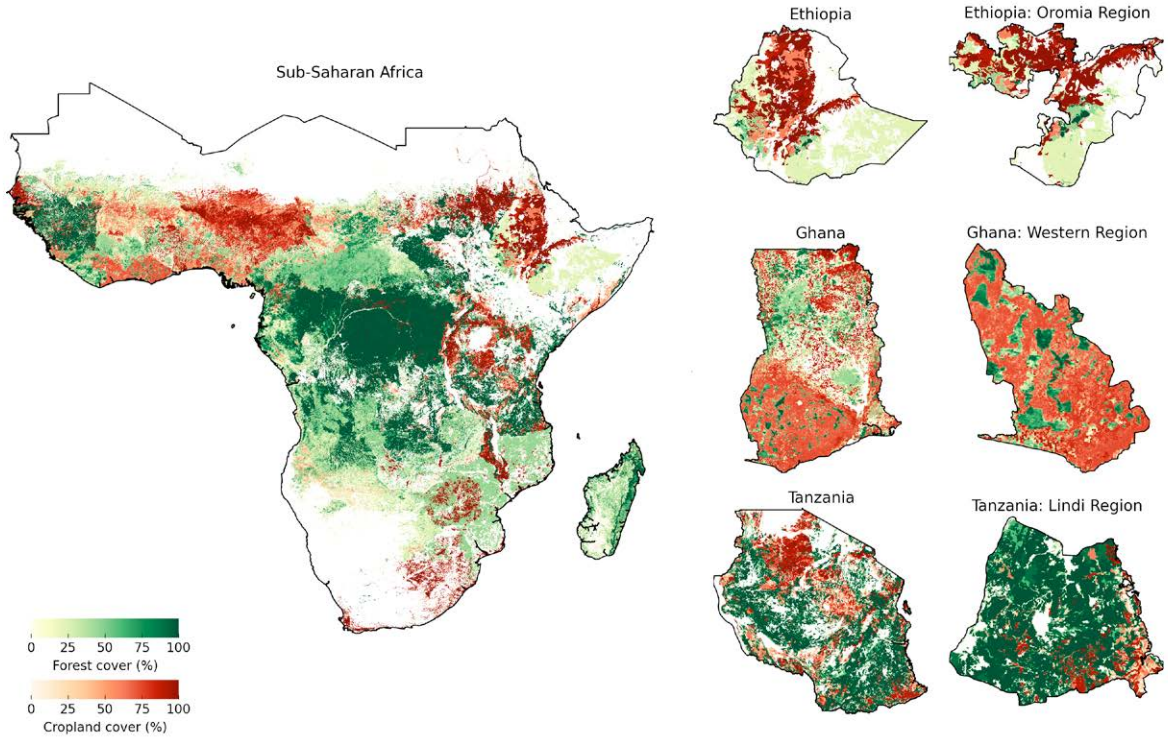
Representation of forest and cropland

Forest vegetation is represented by forest variants 1–4 and woodland classes 9–10. Mosaics of forest and cropland/savannah are also present in classes 7, 8 and 19. Cropland cover includes classes 18, 20 and 21, and part of classes 7 and 19.

Accuracy assessment

The dataset creators assess map accuracy in Africa against other remote sensing data products (eg MODIS VCF), against a continental-scale terrestrial ecoregion map (Olson *et al.* 2001) and with visual comparisons against Landsat data. The authors remark that the data product appears to be of high quality, though a full quantitative analysis of accuracy for the African land cover map does not appear to have been performed. A validation of the global GLC-2000 land cover map measured an overall classification accuracy of 68.6% (Mayaux *et al.* 2006).

GLC-Share



Description

Global Land Cover-SHARE (GLC-SHARE) is a new FAO land cover dataset which combines previously scattered global and cover data in a single data product. It claims to be the "most-reliable global view of planetary land cover assembled to-date". GLC-SHARE complies data from multiple sources. The data used in a given location is a combination of the 'best available' high resolution regional, national and global land cover databases.

Thematic and spatial resolution

The spatial resolution of GLC-SHARE is 30 arcseconds (~1km). GLC-SHARE presents the proportion of each pixel covered by 11 generic land cover types. This land cover classification is advantageous in adhering to internationally agreed standards from the land cover classification system (LCCS) version three, providing consistency and the ready capacity to compare results to other LCCS conforming land cover products. Classes in GLC-SHARE are:

VALUE	DESCRIPTION	PFOREST	PCROP
01	Artificial surfaces	0	0
02	Cropland	0	1
03	Grassland	0	0
04	Tree covered area	1	0
05	Shrubs covered area	0	0
06	Herbaceous vegetation	0	0
07	Mangroves	0	0
08	Sparse vegetation	0	0
09	Baresoil	0	0
10	Snow and glaciers	0	0
11	Water bodies	0	0

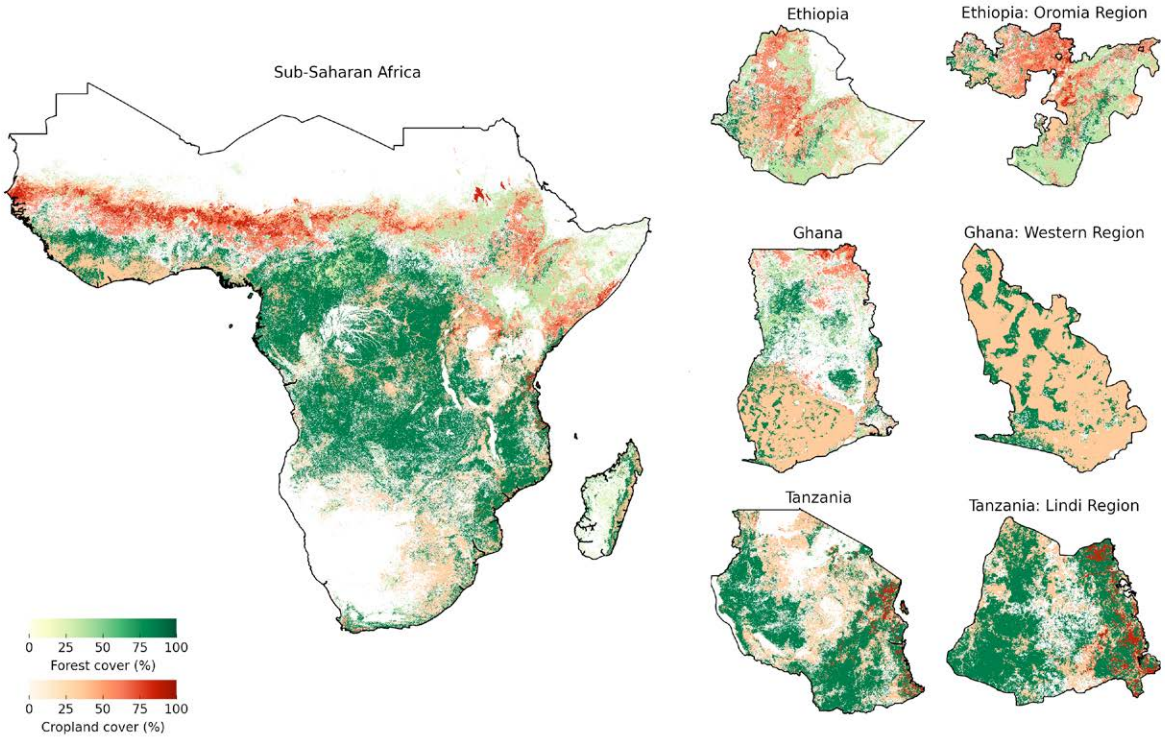
Representation of forest and cropland

Forest cover is represented by 'tree covered area' which is any geographic area dominated by natural tree plants with a cover of 10% or more (following the FAO definition of forest). Tree covered area includes trees planted for afforestation purposes but excludes tree crops in plantations. Cropland is represented by a single land cover class. The cropland land cover class includes herbaceous crops, woody crops (ie plantations and orchards) and multiple or layered crops (eg herbaceous crops planted amongst natural trees).

Accuracy assessment

The producers of GLC-SHARE report accurate classification of 80.2% of pixels. High resolution datasets are available for much of Eastern Africa, whereas Western Africa is only covered by global-scale datasets. Global-scale datasets from GlobCover 2009, MODIS 2010 and the IISA-IFPRI cropland database are mainly used where no local-scale data are available. Where national forest and crop area inventory data are used there are apparent discontinuities between countries. These will relate to the different data sources for different countries and their dissimilar methods. As a very new product no external assessments or comparisons have been conducted.

GlobCover



Description

GlobCover is a widely-used global land cover map generated by the ESA. The GlobCover land cover product uses optical remote sensing data from the MERIS sensor on the now inactive Envisat satellite to classify land cover. Land cover classification was performed largely with unsupervised classification techniques. Global land cover maps have been generated for the years 2005 and 2009; the two maps use different sets of training data which are not recommended for use in measuring land cover change.

Thematic, spatial and temporal resolution

GlobCover has 300m spatial resolution with land cover maps available for 2005 and 2009. A series of 22 generic land cover classes are used which are:

VALUE	DESCRIPTION	PFOREST	PCROP
11	Post-flooding or irrigated croplands (or aquatic)	0	0.7 – 1
14	Rainfed croplands	0	0.7 – 1
20	Mosaic cropland (50–70%) / vegetation (grassland/shrubland/forest) (20–50%)	0 – 0.5	0.5 – 0.7
30	Mosaic vegetation (grassland/shrubland/forest) (50–70%) / cropland (20–50%)	0 – 0.7	0.2 – 0.5
40	Closed to open (>15%) broadleaved evergreen or semi-deciduous forest (>5m)	0.7 – 1	0
50	Closed (>40%) broadleaved deciduous forest (>5m)	0.7 – 1	0
60	Open (15–40%) broadleaved deciduous forest/woodland (>5m)	0.7 – 1	0
70	Closed (>40%) needleleaved evergreen forest (>5m)	0.7 – 1	0
90	Open (15–40%) needleleaved deciduous or evergreen forest (>5m)	0.7 – 1	0
100	Closed to open (>15%) mixed broadleaved and needleleaved forest (>5m)	0.7 – 1	0
110	Mosaic forest or shrubland (50–70%) / grassland (20–50%)	0 – 0.7	0
120	Mosaic grassland (50–70%) / forest or shrubland (20–50%)	0 – 0.5	0
130	Closed to open (>15%) (broadleaved or needleleaved, evergreen or deciduous) shrubland (<5m)	0	0
140	Closed to open (>15%) herbaceous vegetation (grassland, savannas or lichens/mosses)	0	0
150	Sparse (<15%) vegetation	0	0
160	Closed to open (>15%) broadleaved forest regularly flooded (semi-permanently or temporarily) - Fresh or brackish water	0	0
170	Closed (>40%) broadleaved forest or shrubland permanently flooded - Saline or brackish water	0	0
180	Closed to open (>15%) grassland or woody vegetation on regularly flooded or waterlogged soil - Fresh, brackish or saline water	0	0
190	Artificial surfaces and associated areas (Urban areas >50%)	0	0
200	Bare areas	0	0
210	Water bodies	0	0
220	Permanent snow and ice	0	0

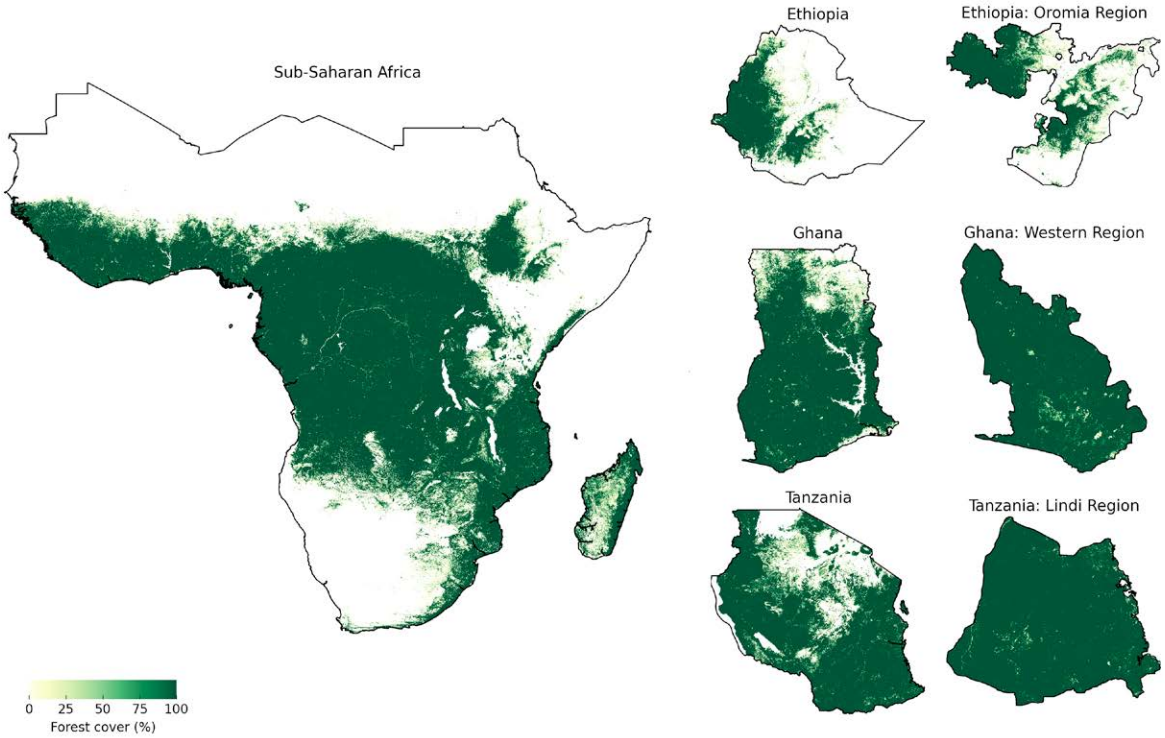
Representation of forest and cropland

Forests in GlobCover are represented by classes 40–100 as well as a proportion of classes 20, 30, 110 and 120. Forest cover is recorded where vegetation is taller than 5m. Cropland cover includes both irrigated (class 11) and rainfed (class 14) agriculture in addition to mosaic classes 20 and 30.

Accuracy assessment

Thematic accuracy is assessed by the dataset creators with a validation database generated by a series of 14 experts (Bontemps *et al.* 2011). An overall accuracy of 58.0% is reported, which takes into account classes that are in agreement (eg class 30 is in agreement with classes 40, 50, 60, 70, 80, 90, 100, 130 and 140). Reported accuracy rises to 65.2% where the reported relative dominance of land cover types is not considered. Class confusions are noted between croplands and grasslands, and between the different varieties of forests.

Hansen



Description

The Hansen map is a high resolution global map of 21st century deforestation events (Hansen *et al.* 2013). In addition to estimates of deforested area, the data product quantifies tree cover in the year 2000 which can be used to calculate annual estimates of forest cover.

The Hansen deforestation map can be viewed online at https://earthenginepartners.appspot.com/science-2013-global-forest

Thematic, spatial and temporal resolution

The Hansen dataset is generated using optical remote sensing reflectance signatures from Landsat which are used to measure proportional tree cover. The Hansen deforestation map provides an estimate of tree cover in the year 2000, annual maps from 2000–2014 of deforestation and a single estimate of tree cover re-growth for the period 2000–2012. Deforestation and re-growth are identified where forest (any level of tree cover) and non-forest (no tree cover) transition. The data have a spatial resolution of 30m.

Representation of forest and cropland

Forests in the Hansen dataset are represented as a proportional tree cover, where trees are defined as vegetation taller than 5m in height. As such, it includes areas of plantations of crops (eg oil palm, rubber) and even some herbaceous crops from forests (Tropek *et al.* 2014). The Hansen dataset can be used to estimate forest area by assuming a threshold of 10% tree cover to separate forest from non-forest. By subtracting the deforested area from forest cover in 2000 and adding 1/12th of re-growth area an annual estimate of tree cover and forest cover change can be generated. We exclude deforestation data from 2013 and 2014 as these layers are not yet included in tree growth estimates. The Hansen dataset does not measure cropland extent.

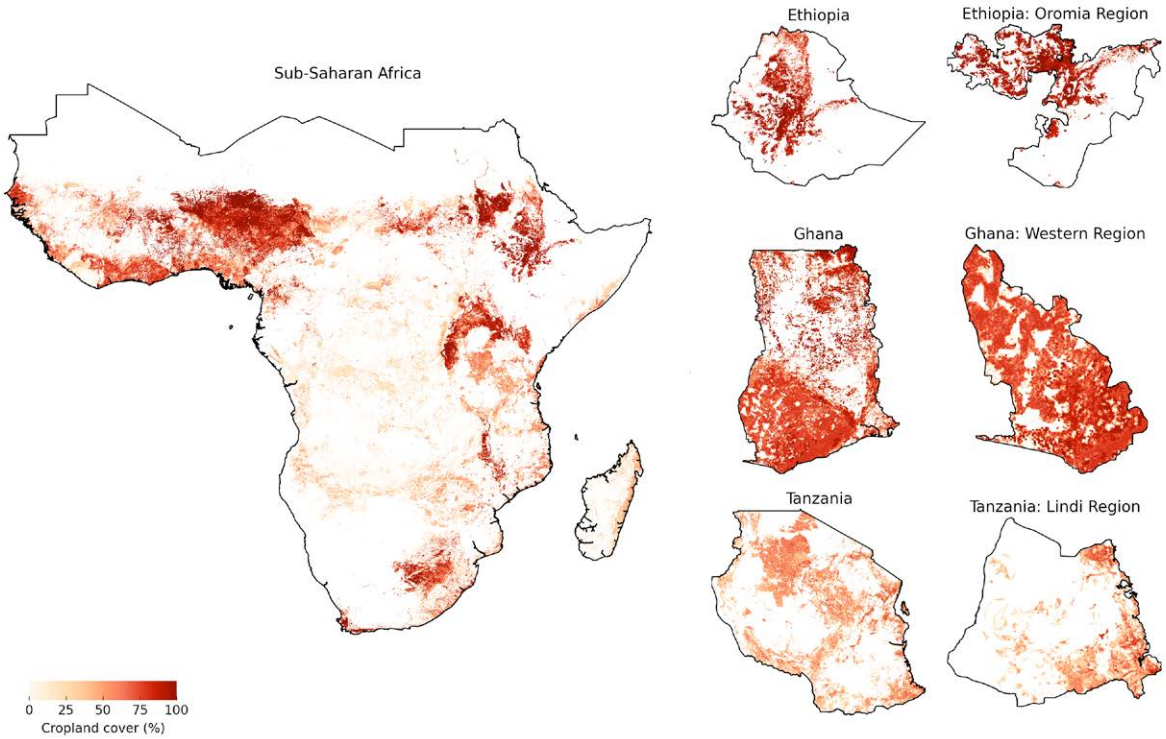
Accuracy assessment

Validation was performed as an independent mapping exercise with accuracies of >80% reported across a range of biomes. The dataset creators identify a possible underestimate of tree cover gain over the study period. Discrepancies are noted between data from FAOSTAT and result of the Hansen deforestation map, particularly in African countries where deforestation rates are underestimated. The definition of forest cover as all vegetation

greater than 5m is a limitation to this product which likely results in a significant overestimate of forest area (Tropek *et al.* 2014).

Qualitative assessment of deforestation events appear to relate well to known areas of deforestation, and the Hansen dataset is a widely-used by researchers as a first-pass tool for identification of areas of deforestation.

IIASA-IFPRI



Description

The International Institute for Applied Systems Analysis (IIASA) International Food Policy Research Institute (IFPRI) dataset is a global cropland map which integrates a range of cropland maps at regional, national and global scales. This map includes data from GlobCover (2005), MODIS LULC, and regional maps from AFRICOVER and national mapping agencies.

Thematic, spatial and temporal resolution

The IIASA-IFPRI dataset is a single representation of cropland area around the year 2005, with a spatial resolution of 30 arcseconds (~1km).

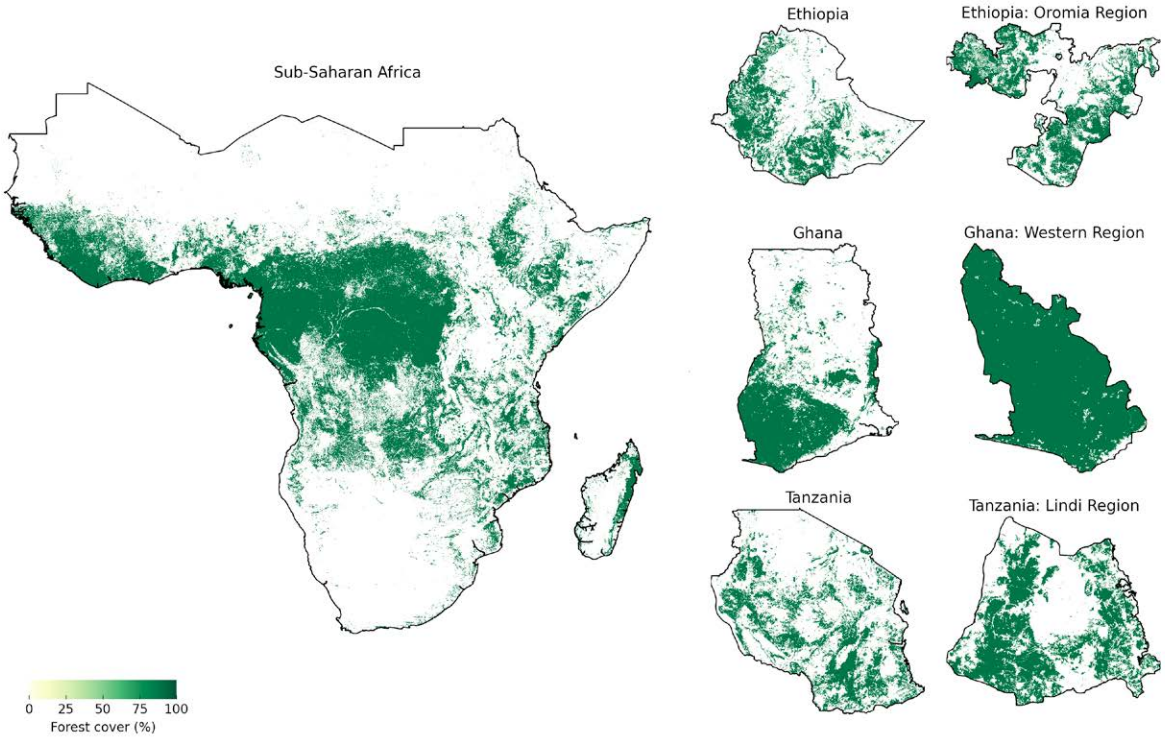
Representation of forest and cropland

Cropland area is represented by the proportion of each pixel that is covered in cropland. No forest area data are given by the IIASA-IFPRI dataset.

Accuracy assessment

The IIASA-IFPRI cropland map is validated with high-resolution imagery through Geo-Wiki, a website that crowdsources cropland identification. An overall accuracy of 82.4% is reported and a lower root-mean-square error (RMSE) reported relative to 'Croplands 2000' when compared against Geo-Wiki validation data.

JAXA



Description

The JAXA forest/non-forest data product is generated using Phased Array type L-band Synthetic Aperture Radar (PALSAR) data from the ALOS. Backscatter data from this radar instrument are used to separate forests (high backscatter) from non-forest (low backscatter) land covers. The data product uses region-dependent backscatter thresholds, as radar backscatter from forests is regionally variable, with the African continent assigned a single backscatter threshold of -15.6 decibels (dB). Mosaics of ALOS PALSAR data were generated for each year, with preference for data showing the minimum response to surface moisture. Where data for a particular year were unavailable (another use for the satellite being disaster monitoring), data were selected from the year before or after.

Thematic, spatial and temporal resolution

The forest/non-forest maps generated by JAXA have a spatial resolution of 25m; as this results in a very large dataset we use the coarser 100m resolution data product here. Data are available for each year from 2007–2010 from ALOS. The launch of ALOS-2 in 2014 will result in additional data in future, though as this is only available in high resolution format we do not consider it here.

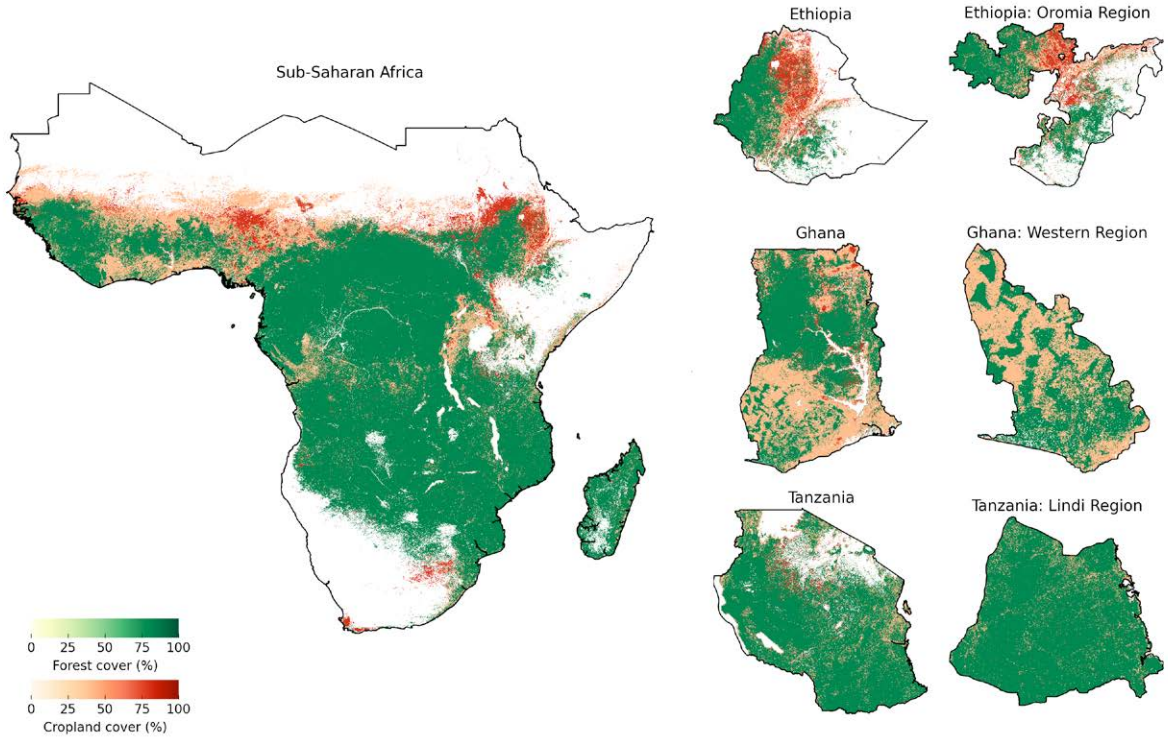
Representation of forest and cropland \

Forest is defined using the FAO definition, requiring natural forest coverage over 0.5 hectares with canopy cover over 10%. The JAXA product will also likely include areas of plantation forestry and orchards which will meet the 10% canopy cover threshold. The JAXA forest/non-forest mask does not measure cropland area.

Accuracy assessment

Mapping accuracy was assessed using Google Earth imagery, resulting in a global classification accuracy of 91% for forests and 85% for non-forests. Accuracy on the African content is reported as 92.8% for forest areas and 93.2% for non-forest areas. Radar backscatter data can experience particular problems in distinguishing high backscatter in forests from soils with a high water content which may result in inter-annual variation in forest area estimates in this data product.

MODIS LULC



Description

The moderate resolution imaging spectrometer (MODIS) land use/land cover change product uses estimates of surface reflectance, land surface temperature and enhanced vegetation index measurements to characterise land cover (Friedl *et al.* 2010). Twelve monthly estimates for each feature are included as inputs allowing the algorithm to exploit information relating to phenology and temporal variability to aid classification. The MODIS LULC data product uses a supervised classification approach where a decision tree algorithm is used to generate class-conditional probabilities for each land cover in each pixel. The classification algorithm is trained using data from 1,860 reference sites distributed across the Earth's land surface.

Prior knowledge about Earth surface characteristics is incorporated into the land cover map to improve predictions. For instance, agriculture is particularly difficult to identify in optical data so the MODIS land cover product merges estimates of cropland vegetation cover from Ramankutty *et al.* (2008) (see 'Croplands 2000') with outputs from the classification. Pixel classifications between years is often unstable, resulting in inter-annual classification changes being common in heterogeneous areas and at ecotones. This is addressed using constraints on year-to-year variation in classification results at each pixel, meaning new pixel labels can only appear where the probability of true change to land cover is high.

Thematic, spatial and temporal resolution

MODIS LULC data have a spatial resolution of 15 arc-seconds (~500m) and are available annually over the period 2001–2012. We exclude the land cover data from 2001 from this analysis, where land cover data are of markedly lower quality. The MODIS LULC data product is available for use with 5 different classification systems; here we use the land cover classes of IGBP which are:

VALUE	DESCRIPTION	PFOREST	PCROP
0	Water	0	0
1	Evergreen Needleleaf forest	0.6 – 1	0
2	Evergreen Broadleaf forest	0.6 – 1	0
3	Deciduous Needleleaf forest	0.6 – 1	0
4	Deciduous Broadleaf forest	0.6 – 1	0
5	Mixed forest	0.6 – 1	0
6	Closed shrublands	0	0
7	Open shrublands	0	0
8	Woody savannas	0.6 – 1	0
9	Savannas	0.6 – 1	0
10	Grasslands	0	0
11	Permanent wetlands	0	0
12	Croplands	0	0.6 – 1
13	Urban and built-up	0	0
14	Cropland/Natural vegetation mosaic	0	0.2 – 0.6
15	Snow and ice	0	0
16	Barren or sparsely vegetated	0	0
254	Unclassified	0	0
255	Fill Value	0	0

Uniquely amongst available data products MODIS LULC has a high enough temporal resolution to estimate deforestation rates and provides an estimate of cropland area. Thus this data product may be of use in identifying which areas of forest that have been lost to cropland expansion.

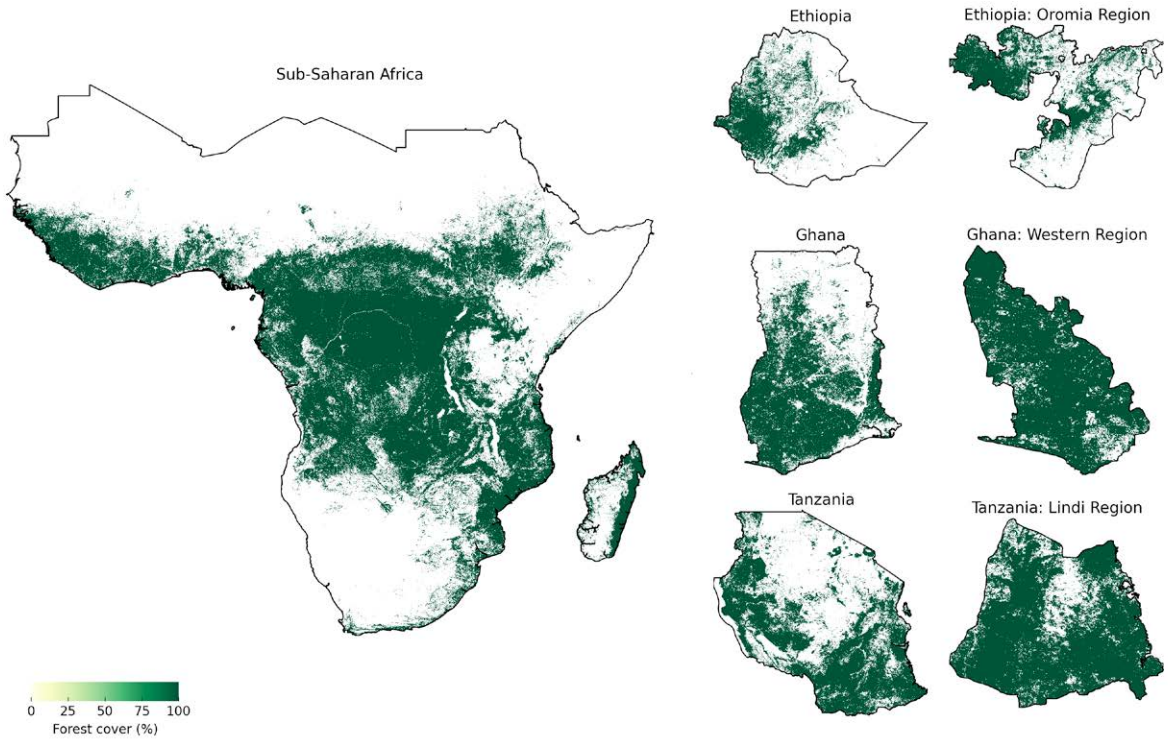
Representation of forest and cropland

Forest in the MODIS LULC data product is represented as evergreen and deciduous broad and needle leaved forest classes, in addition to savannah classes 8 and 9. Croplands are represented by a single cropland class (class 12) in addition to a cropland mosaic class (class 14).

Accuracy assessment

Cross-validation was used to assess classification accuracy, where 10% of training sites are excluded from the classification procedure and used as validation data. The overall accuracy across all classes in the 2005 map is 74.8%. Low classification accuracies were noted in open shublands, woody savannas, savannas, grasslands and agricultural mosaic vegetation classes. Confusion with woody savannahs is substantial, which is frequently misclassified as savanna, forest and cropland mosaic classes. Forest classes generally show good classification accuracy, as does cropland class 12.

MODIS VCF



Description

The MODIS vegetation continuous fields (VCF) product estimates global tree covered density. The MODIS VCF data product uses a range of spectral information from MODIS with a supervised regression tree algorithm to estimate per cent tree cover, a flexible nonlinear model which is robust to vegetation heterogeneity. Training data are generated from 250 Landsat images classified into four classes of per cent tree cover (0, 25, 50, 80+) which were averaged to MODIS 250m spatial resolution to yield a continuous representation of tree cover for that surface. Proportional per pixel estimates of tree cover are an improved thematic representation over discrete classifications with particular advantages in structurally complex landscapes.

Thematic, spatial and temporal resolution

MODIS VCF data are available at 15 arcseconds (~500m) resolution on an annual basis from 2001–2010. We exclude the land cover data from 2001 from this analysis, where land cover data are of markedly lower quality. Data describe proportional tree cover in each pixel ranging from 0–100%.

Representation of forest and cropland

We use a 10% tree cover cut-off to separate forest from non-forest. Tree cover refers to the per cent of skylight obstructed by tree canopies greater than 5m in height. This is not quite the same as crown cover (as used in the FAO definition of forest) which refers to canopy cover plus within-crown skylight. Thus our cut-off should result in a slight under-estimate in forest area relative to the FAO forest definition. The MODIS VCF data product does not estimate cropland area.

Accuracy assessment

The dataset creators have conducted a limited validation using data from Maryland, USA and three sites in Brazil. Their measurements imply an RMSE of ~10% in tree cover estimates and a mean absolute error of ~8.5%. We note substantial inter-annual variations in forest cover estimated by this product when using a 10% tree cover threshold.

Acronyms

AGB	Above ground biomass
ALOS	Advanced land observing satellite
AVHRR	Advanced very high resolution radiometer
CCI	Climate change initiative
ESA	European Space Agency
FAO	Food and Agriculture Organisation
GAEZ	Global agro-ecological zones
GCAD	Global cropland area database
GEDI	Global ecosystem dynamics investigation
IFPRI	International Food Policy Research Institute
IISA	International Institute for Applied Systems Analysis
LCCS	Land cover classification system
LiDAR	Light detection and ranging
LULC	Land use/land cover
MapSPAM	Map spatial production allocation model
MERIS	Medium resolution imaging spectrometer
MMD	Minimum measurable disagreement
MODIS	Moderate resolution imaging spectrometer
NDVI	Normalised difference vegetation index
PALSAR	Phased array type l-band synthetic aperture radar
SSA	Sub-Saharan Africa
USGS	United States Geological Survey
VCF	Vegetation continuous fields

References

Avitabile, V *et al.* (2012) Capabilities and limitations of Landsat and land cover data for above ground woody biomass estimation of Uganda. *Remote Sensing of Environment* 117 366–380.

Avitabile, V *et al.* (2016) An integrated pan-tropical biomass map using multiple reference datasets. *Global change biology* 22(4) 1406–1420.

Baccini, A *et al.* (2012) Estimated carbon dioxide emissions from tropical deforestation improved by carbon-density maps. *Nature Climate Change* 2(3) 182–185.

Bartholomé, E and Belward, A (2005) GLC2000: a new approach to global land cover mapping from Earth observation data. *International Journal of Remote Sensing* 26(9) 1959–1977.

Bontemps, S *et al.* (2011) GLOBCOVER 2009 Products Description Manual Version 1.0. See http://due.esrin.esa.int/page_globcover.php

Bontemps, S *et al.* 'Consistent global land cover maps for climate modelling communities: current achievements of the ESA's land cover CCI.' Presentation at Proceedings of the ESA Living Planet Symposium, Edinburgh, 9–13 September 2013.

Di Gregorio, A and Jansen, L J 'Land cover classification system.' Presentation at Proceedings of the 1st Earth Observation and Environmental Information Conference, Alexandria, Egypt, 13–16 October 1997.

Dubayah, R *et al.* 'The global ecosystem dynamics investigation.' Presentation at the American Geophysical Union, Fall Meeting 2014, San Francisco, 15 December 2014.

FAO (2016) Land cover classification system: Classification concepts and user manual, version 3. www.fao.org/publications/card/en/c/93c7a130-2f12-4f7a-a317-011eef91c5af/

Fisher, B (2010) African exception to drivers of deforestation. *Nature Geoscience* 3(6) 375–376.

Fischer, G *et al.* 'Global agro-ecological zones assessment for agriculture (GAEZ 2008).' Presented at IIASA, Laxenburg, Austria and FAO, Rome, Italy.

Friedl, M *et al.* (2002) Global land cover mapping from MODIS: algorithms and early results. *Remote Sensing of Environment* 83(1) 287–302.

Friedl, M *et al.* (2010) MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote Sensing of Environment* 114(1) 168–182.

Fritz, S *et al.* (2011) Highlighting continued uncertainty in global land cover maps for the user community. *Environmental Research Letters* 6(4) 044005.

Fritz, S *et al.* (2015) Mapping global cropland and field size. *Global change biology* 21(5) 1980–1992.

Geist, H J and Lambin, E F (2002) Proximate Causes and Underlying Driving Forces of Tropical Deforestation. *BioScience* 52(2) 143–150.

Hansen, M *et al.* (2003) Global per cent tree cover at a spatial resolution of 500 meters: First results of the MODIS vegetation continuous fields algorithm. *Earth Interactions* 7(10) 1–15.

Hansen, M *et al.* (2013) High-resolution global maps of 21st-century forest cover change. *Science* 342(6160) 850–853.

Herold, M *et al.* (2012) Building saliency, legitimacy, and credibility towards operational global land cover observations. Remote Sensing of Land Use and Land Cover, Principles and Applications. CRC Press, Taylor Francis, Boca Raton.

Houghton, R A *et al.* (2012) Carbon emissions from land use and land-cover change. *Biogeosciences* 9(12) 5125–5142.

Latham, J (2001) AFRICOVER east Africa. LUCC Newsletter 7 15–16.

Latham, J *et al.* (2014) FAO Global Land Cover (GLC-SHARE) Beta-Release 1.0 Database, FAO Land and Water Division. Data is available at www.glcn.org/index_en.jsp

Le Toan, T *et al.* (2011) The BIOMASS mission: Mapping global forest biomass to better understand the terrestrial carbon cycle. *Remote sensing of environment* 115(11) 2850–2860.

Lund, H G (2008) Definitions of forest, deforestation, afforestation, and reforestation. Forest Information Services. Gainesville, VA

Mayaux, P *et al.* (2004) A new land-cover map of Africa for the year 2000. *Journal of Biogeography* 31(6) 861–877.

Mayaux, P *et al.* (2006) Validation of the global land cover 2000 map. *IEEE Transactions on Geoscience and Remote Sensing* 44(7) 1728–1739.

Mitchard, E T *et al.* (2013) Uncertainty in the spatial distribution of tropical forest biomass: a comparison of pan-tropical maps. *Carbon balance and management* 8(1) 1.

Olson, D M *et al.* (2001) Terrestrial Ecoregions of the World: A New Map of Life on Earth. A new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. *BioScience* 51(11) 933–938.

Pittman, K *et al.* (2010) Estimating global cropland extent with multi-year MODIS data. *Remote Sensing* 2(7) 1844–1863.

Prestele, R *et al.* (2016) Hotspots of uncertainty in land-use and land-cover change projections: a global-scale model comparison. *Global change biology* 22(12) 3967–3983.

Ramankutty, N *et al.* (2008) Farming the planet: 1. Geographic distribution of global agricultural lands in the year 2000. *Global Biogeochemical Cycles* 22(1).

Ramankutty, N and Foley, J A (1998) Characterizing patterns of global land use: An analysis of global croplands data. *Global Biogeochemical Cycles* 12(4) 667–685.

Saatchi, S *et al.* (2011) Benchmark map of forest carbon stocks in tropical regions across three continents. *Proceedings of the National Academy of Sciences* 108(24) 9899–9904.

Shimada, M *et al.* (2014) New global forest/non-forest maps from ALOS PALSAR data (2007–2010). *Remote Sensing of Environment* 155 13–31.

Smith, R (2010) Global Forest Resources Assessment 2000. Main report, FAO Forestry Paper 163, UNFAO.

Teluguntla, P G *et al.* (2015) Global Cropland Area Database (GCAD) Derived from Remote Sensing in Support of Food Security in the Twenty-First Century: Current Achievements and Future Possibilities, Taylor & Francis: Boca Raton, FL, USA.

Thenkabail, P S *et al.* (2009) Global irrigated area map (GIAM), derived from remote sensing, for the end of the last millennium. *International Journal of Remote Sensing* 30(14) 3679–3733.

Tropek, R *et al.* (2014) Comment on "High-resolution global maps of 21st-century forest cover change". *Science* 344(6187) 981–981.

You, L *et al.* (2014) Generating global crop distribution maps: from census to grid. *Agricultural Systems* 127 53–60.

Yu, L *et al.* (2013) FROM-GC: 30m global cropland extent derived through multisource data integration. *International Journal of Digital Earth* 6(6) 521–533.

Related reading

Hou Jones, X and Franks, F (2015) Food vs forests in sub-Saharan Africa: a challenge for the SDGs. IIED Briefing. IIED, London. See http://pubs.iied.org/17322IIED/

IIED, Food demand and forests in sub-Saharan Africa. www.iied.org/food-demand-forests-sub-saharan-africa

Effective monitoring of deforestation and cropland expansion in Africa requires reliable estimates of land cover area. However, continental scale land cover datasets generated solely or partially through remote sensing technologies show large differences in the extent and spatial distribution of forest and cropland. In this working paper we identify 14 of the most commonly used land cover data products, summarise their spatial, temporal and thematic properties, and compare their estimates of forest and cropland area. We describe the limitations to land cover data that result from their divergent predictions and discuss the implications for using this data to assess the agricultural drivers of deforestation in Africa.

IIED is a policy and action research organisation. We promote sustainable development to improve livelihoods and protect the environments on which these livelihoods are built. We specialise in linking local priorities to global challenges. IIED is based in London and works in Africa, Asia, Latin America, the Middle East and the Pacific, with some of the world's most vulnerable people. We work with them to strengthen their voice in the decision-making arenas that affect them – from village councils to international conventions.

iied

International Institute for Environment and Development
80-86 Gray's Inn Road, London WC1X 8NH, UK
Tel: +44 (0)20 3463 7399
Fax: +44 (0)20 3514 9055
email: info@iied.org
www.iied.org

Knowledge Products

Funded by:

UKaid from the British people

the David & Lucile Packard FOUNDATION

This research was funded by UK aid from the UK Government, with support for the Ethiopia component of the project from The Packard Foundation. However, the views expressed do not necessarily reflect the views of the UK Government or The Packard Foundation.