ISPRS Journal of Photogrammetry and Remote Sensing 126 (2017) 225–244

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Automated cropland mapping of continental Africa using Google Earth Engine cloud computing

Jun Xiong [a,c,*], Prasad S. Thenkabail [a], Murali K. Gumma [b], Pardhasaradhi Teluguntla [a,c], Justin Poehnelt [a], Russell G. Congalton [d], Kamini Yadav [d], David Thau [e]

[a] U.S. Geological Survey (USGS), 2255, N. Gemini Drive, Flagstaff, AZ 86001, USA
[b] International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, Hyderabad, India
[c] Bay Area Environmental Research Institute (BAERI), 596 1st St West Sonoma, CA 95476, USA
[d] University of New Hampshire, NH, USA
[e] Google, MountainView, CA, USA

ARTICLE INFO

Article history:
Received 11 July 2016
Received in revised form 9 December 2016
Accepted 24 January 2017

Keywords:
Cropland mapping
Classification
MODIS
Remote sensing products
Google Earth Engine
Africa
Automated cropland mapping algorithm

ABSTRACT

The automation of agricultural mapping using satellite-derived remotely sensed data remains a challenge in Africa because of the heterogeneous and fragmental landscape, complex crop cycles, and limited access to local knowledge. Currently, consistent, continent-wide routine cropland mapping of Africa does not exist, with most studies focused either on certain portions of the continent or at most a one-time effort at mapping the continent at coarse resolution remote sensing. In this research, we addressed these limitations by applying an automated cropland mapping algorithm (ACMA) that captures extensive knowledge on the croplands of Africa available through: (a) ground-based training samples, (b) very high (submeter to five-meter) resolution imagery (VHRI), and (c) local knowledge captured during field visits and/or sourced from country reports and literature. The study used 16-day time-series of Moderate Resolution Imaging Spectroradiometer (MODIS) normalized difference vegetation index (NDVI) composited data at 250-m resolution for the entire African continent. Based on these data, the study first produced accurate reference cropland layers or RCLs (cropland extent/areas, irrigation *versus* rainfed, cropping intensities, crop dominance, and croplands *versus* cropland fallows) for the year 2014 that provided an overall accuracy of around 90% for crop extent in different agro-ecological zones (AEZs). The RCLs for the year 2014 (RCL2014) were then used in the development of the ACMA algorithm to create ACMA-derived cropland layers for 2014 (ACL2014). ACL2014 when compared pixel-by-pixel with the RCL2014 had an overall similarity greater than 95%. Based on the ACL2014, the African continent had 296 Mha of net cropland areas (260 Mha cultivated plus 36 Mha fallows) and 330 Mha of gross cropland areas. Of the 260 Mha of net cropland areas cultivated during 2014, 90.6% (236 Mha) was rainfed and just 9.4% (24 Mha) was irrigated. Africa has about 15% of the world's population, but only about 6% of world's irrigation. Net cropland area distribution was 95 Mha during season 1, 117 Mha during season 2, and 84 Mha continuous. About 58% of the rainfed and 39% of the irrigated were single crops (net cropland area without cropland fallows) cropped during either season 1 (January-May) or season 2 (June-September). The ACMA algorithm was deployed on Google Earth Engine (GEE) cloud computing platform and applied on MODIS time-series data from 2003 through 2014 to obtain ACMA-derived cropland layers for these years (ACL2003 to ACL2014). The results indicated that over these twelve years, on average: (a) croplands increased by 1 Mha/yr, and (b) cropland fallows decreased by 1 Mha/year. Cropland areas computed from ACL2014 for the 55 African countries were largely underestimated when compared with an independent source of census-based cropland data, with a root-mean-square error (RMSE) of 3.5 Mha. ACMA demonstrated the ability to hind-cast (past years), now-cast (present year), and forecast (future years) cropland products using MODIS 250-m time-series data rapidly, but currently, insufficient reference data exist to rigorously report trends from these results.

Abbreviations: AVHRR, advanced very high resolution radiometer; FAO, Food and Agricultural Organization of the United Nations; FROMGC, 30 m global cropland extent derived through multisource data; GCEV1, global cropland extent version 1; GFSAD250, global food security support analysis data (GFSAD) cropland products of Africa at 250-m resolution; GLC2000, global land cover for the nominal year 2000; GRIPC, global rainfed, irrigated, and paddy croplands; LULC2000, land use land cover for the nominal year 2000; MCD12Q1, MODIS land cover type product; MERIS, MEdium Resolution Imaging Spectrometer; MODIS, Moderate Resolution Imaging Spectroradiometer; SPOT, Satellite Pour l'Observation de la Terre.

* Corresponding author at: U.S. Geological Survey (USGS), 2255, N. Gemini Drive, Flagstaff, AZ 86001, USA.
E-mail address: jxiong@usgs.gov (J. Xiong).

http://dx.doi.org/10.1016/j.isprsjprs.2017.01.019
0924-2716/© 2017 The Authors. Published by Elsevier B.V. on behalf of International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS).
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

© 2017 The Authors. Published by Elsevier B.V. on behalf of International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS). This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

The extent, distribution, and characteristics (e.g., irrigation *versus* rainfed, cropping intensity, crop types) of croplands are factors that have long been identified as fundamental influences on agricultural development pathways, food security scenarios, and poverty reduction (Jayne et al., 2014). Estimates show that 52% of the world's remaining arable land is in Africa, yet most of this land is concentrated in just eight countries (Algeria, Democratic Republic of the Congo, Ethiopia, Morocco, Nigeria, South Africa, Sudan, Uganda), while a number of the remaining countries contain large rural populations clustered in remarkably small areas (Chamberlin et al., 2014). Demography of Africa is projected to change exponentially, where the population is expected to increase from the current 1.2 billion to nearly 4 billion by the end of the century (Gerland et al., 2014). A quarter of the population is undernourished and many countries experience famines in sub-Saharan Africa (Clover, 2010). In this context, timely and dependable information on agricultural croplands of Africa is a prerequisite necessity to (i) isolate the agricultural croplands to assess crop water use, crop productivity, and crop water productivity, and (ii) investigate how the croplands respond to different climatic conditions (Waldner et al., 2015).

Global land use/land cover (LULC) products such as global land cover 2000 (Giri et al., 2005), GlobCover 2005/2009 (Arino et al., 2007), Global Land Cover-SHARE (Latham et al., 2014), and MODIS (Moderate Resolution Imaging Spectroradiometer) Land Cover (Friedl et al., 2002) do have cropland classes. However, to use these products as accurate and reliable cropland estimation for the practical purpose is questionable. For example, Cropland estimates derived from GlobCover are 20% higher than those derived from MODIS globally (Fritz et al., 2011a,b). Further, the spatial location of the croplands between any two of these global LULC products varies substantially. These factors have led to differences in cropland areas between various products which is as much as staggering 300 Mha globally (varying from 1.5 to 1.8 billion hectares). For example, the Food and Agricultural Organization (FAO) of the United Nations (UN) estimates that, around the year 2010, there was 319 Mha of croplands in Africa compared to the significantly lower MODIS land cover and GlobCover estimates of 277 Mha and 152 Mha, respectively. There are many reasons for such differences such as 1. these products are more focused on LULC systems than on agricultural systems, 2. definition issues, 3. resolution of the data used, 4. other data characteristics (e.g., spectral, radiometric), and 5. Methods adopted. Further, in these products croplands are not a single land cover class, but are contained within the mosaic of classes without specific agricultural information such as irrigation, cropping intensity, or crop type. All of these factors lead to substantial uncertainties in cropland assessment and related products of cropland water use and food security assessment and reporting.

Further, there are several cropland studies. Time-series remotely sensed data are established as effective tool in cropland mapping (Esch et al., 2014) and have been successfully implemented at regional-scale (Bégué et al., 2014; Ding et al., 2014; Gumma et al., 2014; Helmholz et al., 2014; Teluguntla et al., 2015a,b) as well as at global scale (Chen et al., 2015; Pittman et al., 2010; Radoux et al., 2014; Salmon et al., 2015; Thenkabail and Wu, 2012; Wang et al., 2015). Various aspects of croplands are mapped such as irrigated areas (Conrad et al., 2016; PeñArancibia et al., 2016; Salmon et al., 2015; Thenkabail and Wu, 2012), rainfed areas (Biradar et al., 2009; Salmon et al., 2015), cropping intensities (Qiu et al., 2014), and crop types (Gumma et al., 2014; Zhang et al., 2015; Zhong et al., 2014; Zhou et al., 2016), and cropland fallows (Müller et al., 2015). There are many methods and techniques adopted for cropland classification that include phenology based algorithms (Dong et al., 2015; Jeganathan et al., 2014; Pan et al., 2015), classification regression trees (Deng and Wu, 2013; Egorov et al., 2015; Ozdogan and Gutman, 2008), decision tree algorithms (Friedl and Brodley, 1997; Shao and Lunetta, 2012), Fourier harmonic analysis (Zhang et al., 2015), spectral matching techniques (Dheeravath et al., 2010), support vector machines (Mountrakis et al., 2011), random forest algorithm (Tatsumi et al., 2015) and a number of other machine learning algorithms (DeFries, 2000; Duro et al., 2012; Lary et al., 2016; Pantazi et al., 2016). Many studies adopted supervised and unsupervised classification approaches. Supervised methods (Egorov et al., 2015) rely extensively on *in situ* data or on human interpretation of spectral signatures, making the classification process resource-intensive, time-consuming, and difficult to repeat over space and time (Zhong et al., 2014). So, when rich sets of *in situ* data are lacking, as is often the case in Africa, supervised approaches lead to uncertainties. Unsupervised approaches require far less *in situ* data or human interpretation but they require large volumes of *in situ* data for class identification and validation data.

Specific to continental Africa, amongst existing cropland products there has been large disagreement (Fritz and See, 2008; Giri et al., 2005; Hansen and Reed, 2010; Herold et al., 2008; McCallum et al., 2006) especially in the extent of the cultivated areas and their spatial distribution (Fritz et al., 2011a; Salmon et al., 2015; Teluguntla et al., 2015a,b; Thenkabail and Wu, 2012; Waldner et al., 2015) as a result of fragmented and heterogeneous rural landscapes (Lobell and Asner, 2004) and low agricultural intensification (Pittman et al., 2010) throughout continental Africa. The challenges of mapping cropland in Africa also include: (a) spatial structure of the agricultural landscape (Vancutsem et al., 2012), (b) spectral similarity with grassland, mainly in arid and semi-arid areas (Herold et al., 2006; McCallum et al., 2006), (c) high regional variability in terms of agricultural systems and calendars between the hyper-arid Sahara and other agro-ecological zones (Vintrou et al., 2012).

Further, the current state-of-art using the above methods and approaches is mostly limited to producing cropland products for a given period, or for a growing season, or for a particular year. However, such a process over very large areas such as continent will always have limitations in availability of extensive collection of reference data. The biggest difficulty in cropland mapping is in the lack of algorithms that accurately reproduce cropland products year after year or season after season. So, more recently, there are efforts at producing cropland products by developing automated algorithms (Jamali et al., 2014; Waldner et al., 2015; Yan and Roy, 2014). Thenkabail et al. developed rule-based ensemble decision-tree Automated Cropland Classification algorithms (ACCA's) to produce cropland *versus* non-croplands across years for Australia, Tajikistan and California (Teluguntla et al., 2016; Thenkabail and Wu, 2012; Wu et al., 2014). Waldner et al. (2015) used a baseline map generated from five knowledge-based temporal features to train an automated support vector machines (SVM) classifier on selected areas in Argentina, Belgium, Ukraine, and

China. However, these automated algorithms are currently applied only to small pilot studies and not over large areal extent such as the African Continent.

Given the above discussions, the overarching goal of this research was to develop and test automated cropland mapping algorithms (ACMAs) over a very large area such as a continent with an ability to automatically and accurately reproduce cropland products year after year and season after season using MODIS 250-m 16-day time-series data. Africa was chosen given its importance for food security in the twenty-first century and to address the challenge of mapping complex agricultural systems. The spatial, temporal, and spectral specifications of MODIS are considered as highly suitable for land use and land cover (LULC) classifications, especially for cropland extent and area mapping (Hentze et al., 2016). The Google Earth Engine (GEE) cloud computing platform was used in this project to generate the products as well as collecting reference data. The GEE is a system designed to enable petabyte-scale, scientific analysis and visualization of geospatial datasets. Earth Engine provides a consolidated environment including a massive data catalog co-located with thousands of computers for analysis. The user-friendly front-end provides a workbench environment to allow interactive data and algorithm development, and support for in-the-field activities such as validation, ground-sampling, and crowd-sourcing. We first, develop ensemble decision-tree algorithm ACMA for the year 2014 for the African continent, and then tested and validated ACMA for the same year. This was followed by validation of ACMA algorithm for 11 independent years (2003–2013). Finally, we deploy the ACMA algorithm on Google Earth Engine (GEE) cloud computing platform, so scientists and practitioners can routinely reproduce cropland products of Africa year after year.

2. Data

2.1. Study area

The study area included the entire African continent which extends from approximately 38°N to 35°S latitude, occupies 30.3 million km^2, and has several distinct geologic and biogeographic regions with varying land cover types. For example, Sahara, the largest hot desert in the world, comprises much of the land found within North Africa, excluding the fertile coastal region situated against the Mediterranean Sea, the Atlas Mountains of the Maghreb, and the Nile Valley of Egypt and Sudan. Savannas, or grasslands, cover almost half of Africa, more than 13 million km^2. These grasslands make up most of central Africa, beginning south of the Sahara and the Sahel and ending north of the continents southern tip. Also, 80% of Africa's rain forest is concentrated in central Africa, along the Congo River basin. Swahili Coast, stretches about 1610 km along the Indian Ocean, from Somalia to Mozambique, where vegetated areas are located on a narrow strip just inland from the coastal sands and heavy cultivation has diminished the diversity of plant species in this interior area. Southern Africa will be one of the regions in the world whose crop production is most affected by climate change such as higher temperatures and reduced water supplies, along with other factors like biodiversity loss and ecosystems degradation (Lobell et al., 2008). All the raster and vector data in entire Africa continent were produced in Geographic projection (WGS84) at a spatial resolution of 0.0022458 degrees (equivalent to 250 m at the equator).

The FAO Global Agro-Ecological Zones were used as zoning basis (FAO et al., 2012). The entire Africa was divided into eight major agro-ecological zones (Fig. 1) based on climate, soils, and terrain data that in turn indicates the length of crop growing period.

2.2. Existing cropland/LULC reference maps

Available land use/land cover (LULC) reference maps of Africa (Table 1) from different sources vary widely in how they are defined, derived, and mapped using a wide range of data, and methods and have different projections, formats, resolutions, and LULC categories. Even though they are used widely in LULC research, inconsistencies and uncertainties make their use as reliable baseline maps questionable (Teluguntla et al., 2015a,b). Thereby, we expanded our investigation of the cropland extent by incorporating these studies into a comprehensive baseline crop layer (Fig. 3) in Section 3.2.

2.3. Reference samples repository

In-situ samples collected from ground data is always the first step to establishing knowledge for the classifier in classification. These *in situ* data are supposed to provide the most accurate information by definition. However, they are often not an ideal gold standard but degraded by error (Foody, 2010) because of small samples size, sampling bias and inconsistent labeling. A web-based system for supporting classification have been used in the past for general land cover (Fritz et al., 2009; Tsendbazar et al., 2015). In this paper, web-based data was developed (in addition to extensive ground data) using multiple sources and consolidated on the GEE platform.

The reference samples repository consists of following components: The project developed a ground data-collection mobile app https://croplands.org/mobile that can be downloaded and run on a smart phone device. This mobile app allows users to collect geo-references ground data that includes digital photos, cropland data required for the project (cropland *versus* cropland fallows, irrigated *versus* rainfed, cropping intensity). All data so collected from anywhere by anyone in the world is automatically uploaded to the project server. All data so collected from anywhere in the world is automatically uploaded to the project server. All data samples, so collected are further reviewed in the online image-interpretation tool (https://croplands.org/app/data/classify) to ensure that the samples are centered on the farm field using sub-meter to 5-meter very high-resolution imagery (VHRI) data from sensors such as Worldview 2, QuickBird, and IKONOS. Reference ground data for Africa were collected through several field campaigns by the project team in May, June, and August 2014 to coincide with the peak cropping seasons in different parts of Africa. Field information was collected from 250 m × 250 m homogeneous plots. A total of 1381 samples were collected from Ethiopia, Ghana, Kenya, Malawi, Mali and Uganda. Reference data was also collected from several other sources. First, some other global/region projects (Tateishi et al., 2014; Zhao et al., 2014) shared with us valuable reference datasets. To incorporate them into our project, we converted their labeling system to be consistent with the labeling scheme of our project. Second, ~500 reference cropland samples were selected from a series of published literature for selected areas of Africa based on detailed studies using VHRI or high-resolution imagery such as Landsat (Haack et al., 2014; Kidane et al., 2012; Rembold et al., 2000; Shalaby and Tateishi, 2007; Were et al., 2013; Zucca et al., 2015).

Overall, there were total 3265 reference samples (Fig. 1) spread across the eight consolidated agro-ecological zones (AEZs) of the African continent. Of these 953 reference samples were collected during the field visit by the team and the rest 2312 reference samples were sourced from partners/collaborators (Fig. 1). When the full reference samples repository was established through above approaches, every sample was then marked into "training" and "validation" groups. A random 70–30% splitting of the 3265 were used to separate 2285 samples for "training" and the rest 953 for

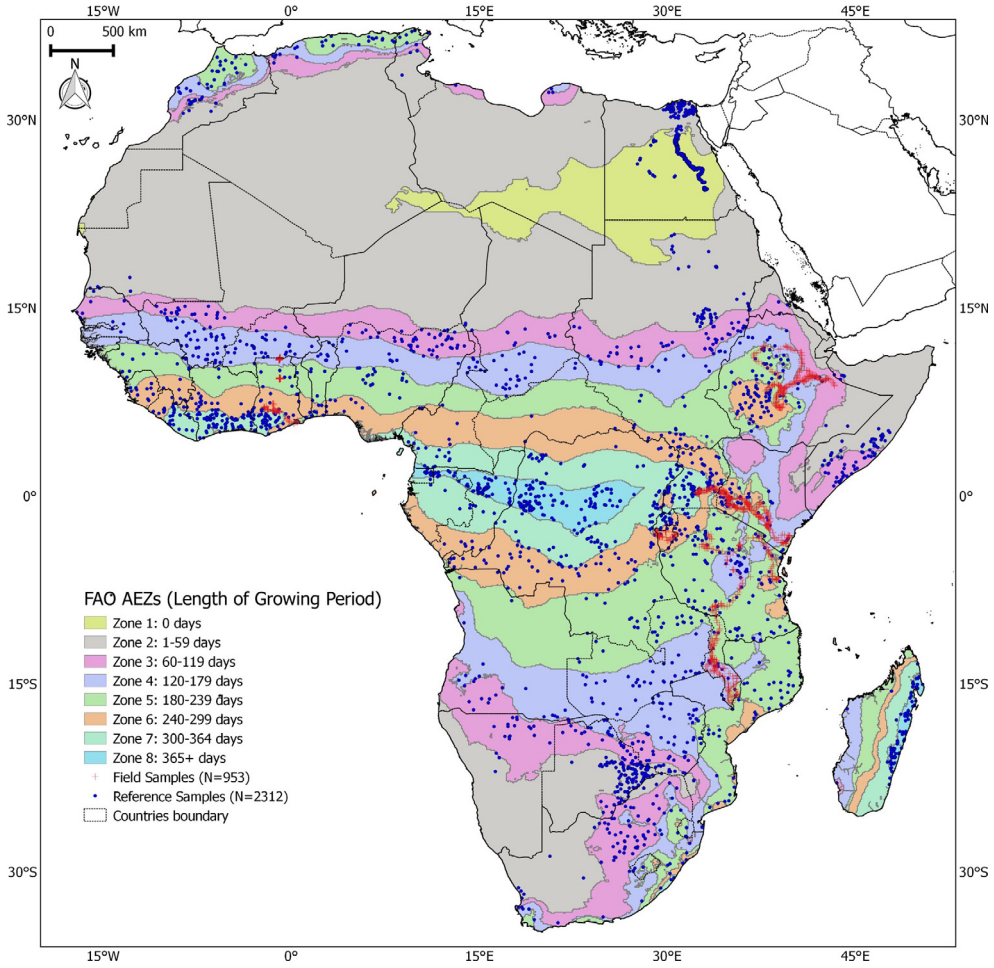

Fig. 1. The United Nations (UN) Food and Agriculture Organization (FAO) Global Agro-Ecological Zones (AEZs) and distribution of reference samples repository in Africa continent. [Note: Initial 15 AEZs were consolidated to final 8 to eliminate AEZs with zero or insignificant agriculture such as in the Sahara Desert.]

"validation". The 3265 training samples were used to create knowledge through ideal spectral libraries. Ground data samples repository collected during the field visit includes mostly pure classes. However, there are still a significant number of mixed classes because of heterogeneous landscape. In order to overcome this heterogeneity, we combine the homogeneous and heterogeneous samples and further use MODIS NDVI signatures to determine distinct and separable groups of classes like the ones illustrated in Fig. 4. These distinct class signatures were then used in the algorithm.

These validation datasets are publicly available for download at the following address: https://croplands.org/app/data/search. Also, the independent accuracy assessment team further added additional validation samples that are hidden to mapping.

2.4. MODIS NDVI times series data

The MODIS 250 m 16-day composite NDVI product was found to have high temporal resolutions to overcome the data gap because of cloud cover and harmattan haze during the monsoon season over Africa (Leroux et al., 2014; Vintrou et al., 2012). Hosted on Google Earth Engine (GEE), the MYD13 product is computed from daily atmospherically corrected bi-directional surface reflectance that has been masked for water, clouds, heavy aerosols, and cloud shadows. Google Earth Engine, using state-of-the-art cloud-computing and storage capabilities, has archived a large catalog of earth observation data and enabled the scientific community to work on petabytes of satellite imagery rapidly using parallel processing (Hansen et al., 2013).

In this paper, NDVI time-series spanning the entire year (January–December 2014) was used as a reference year because most of the ground samples and very high spatial resolution imagery (VHRI) collected in the same year and 2014 is a precipitation normal year. The precipitation data used here is from CHIRP (Funk et al., 2014), which is a 30+ year quasi-global rainfall dataset. Spanning 50°S to 50°N (and all longitudes), starting in 1981 to near-present, CHIRPS incorporates 0.05° resolution satellite imagery with *in situ* station data to create gridded rainfall time series for trend analysis and seasonal drought monitoring.

3. Methodology

3.1. Method overview

The proposed methodology is presented in Fig. 2. First, MODIS 16-day 250 NDVI imagery composite of the African continent was stratified by (1) Masking out the non-cropland area using 250 m baseline cropland mask of Africa (Fig. 3), (2) Sub-setting masked area into eight consolidated FAO agro-ecological zones (AEZs) (Fig. 1), and (3) clustering each cropland mask of the 8 AEZs into 25 unique clusters using K-means algorithm for a total of 200 classes. Second, ground samples from reference samples' repository (Section 2.3) were split into training part and validation part. The former was used to characterize unique ideal time-series sig-

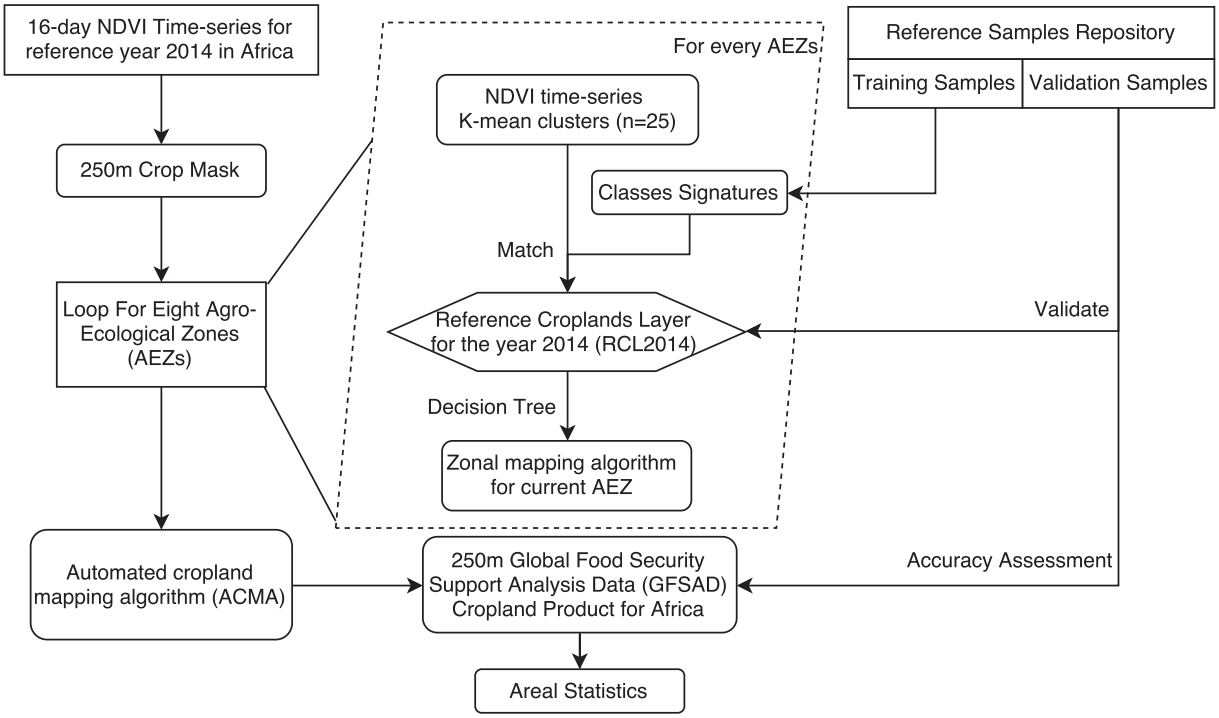

Fig. 2. Schematic diagram of the methodology and area statistics used in creating reference cropland layer for the year 2014 (RCL2014) and the Automated cropland mapping algorithm (ACMA) for continental Africa.

Table 1
Datasets used in creating 250 m cropland mask of Africa in terms of their reference, data source, resolution, and time interval.

Name	Institution	Sensors	Resolution	Time	Classes	Reference
Globcover	ESA	MERIS	300 m	2005, 2009	LULC	Defourny et al. (2009)
Africover	FAO	Landsat 7	30 m	1995–2002	LULC	Kalensky (1998)
LULC 2000	USGS	AVHRR	2000 m	2000	LULC	Soulard et al. (2014)
GLC 2000	JRC	SPOT	1/112°	2000	LULC	Fritz et al. (2010)
MCD12Q1	NASA	MODIS	500 m	2004 - now	LULC	Leroux et al. (2014)
MODIS-JRC	JRC/MARS	MODIS, Landsat	250 m	2009	LULC	Vancutsem et al. (2012)
GCEV1	USGS	MODIS, Landsat	1000 m	2010	Cropland	Teluguntla et al. (2015a,b)
Global30	NGCC	Landsat 7	30 m	2010	LULC	Chen et al. (2015)
FROMGC	CESS	Landsat 7	30 m	Circa 2010	LULC	Gong et al. (2013)
GRIPC	BU	MODIS	500 m	Circa 2005	Cropland	Salmon et al. (2015)

natures. Third, clustered classes from each of the 8 AEZs generated using k-means algorithm were grouped together through quantitative spectral matching techniques (QSMTs) and the group of similar cluster classes was matched with the ideal spectra to identify and label classes. The class labeling is further verified through ground data, VHRI, field visits, and through external sources. The process lead to an accurate reference cropland layer (RCL) of Africa for the year 2014 (RCL2014). The cropland knowledge available in the RCL2014 was then coded in an ensemble decision-tree automated cropland mapping algorithm (ACMA) to accurately replicate cropland products through Automated cropland layer for the year 2014 (ACL2014). Once this is achieved accurately, ACMA was deployed on the Google Earth Engine to create automated cropland layers for independent years from 2003 to 2013.

3.2. 250-m baseline crop mask

Ten previous LULC products of Africa (Table 1) can be put into two types: LULC map following certain Land Cover Classification Systems, where cropland was labeled as (i) one class (Globcover, LULC 2000, Global30) or (ii) multiple classes (MCD12Q1, FROMGC, GRIPC), or (iii) a cropland layer with different intensity levels as percentage (GCEV1, CUI). For (i), cropland class was recoded to 1 while another non-cropland was masked out; For (ii), cropland classes was recoded to 1, if cropland exists in any other mixture classes, count them in. For (iii), a visual analysis of the products in comparison with on Google Earth imagery then the threshold value was set to make sure most of the pixel contained cropland was labeled.

Then, the following processes were applied to integration the different datasets:

1. Rasterizing: vector datasets (Africover, CUI and SADC) are converted into a 250-m resolution raster file with "mode", which means the feature with the largest area in the cell yields the attribute assigned to the 250 m pixel cell.
2. Reproject & Resampling: Datasets were reprojected to the Geographic projection (WGS84) at a 250-m spatial resolution.
3. Aggregation: All the resampled layers have been aggregated to a single crop mask. "Aggregate" means the pixel was set to "cropland" if any layer tell it is a "crop" pixel and ignore other "non-crop" status.

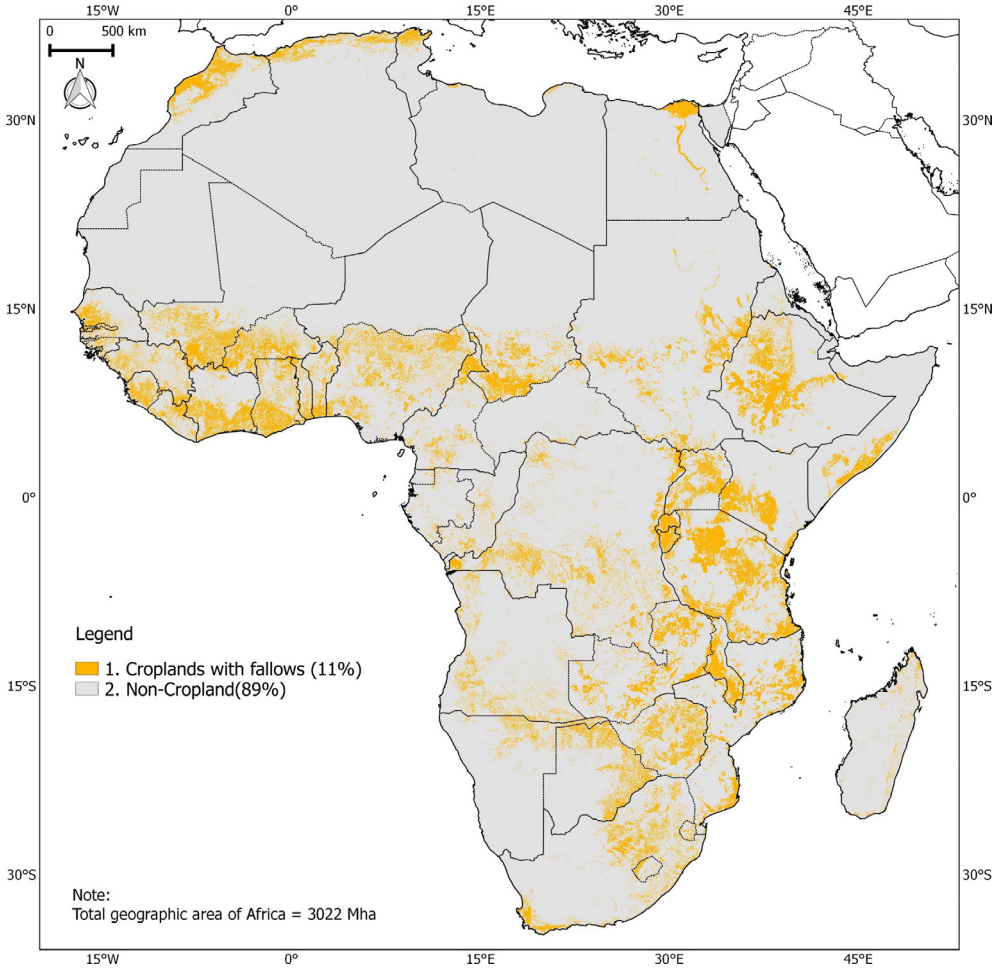

Fig. 3. Baseline cropland mask of Africa based on 10 pathfinding studies. Aggregated 250 m cropland mask derived from 10 previous studies in Africa, including cropland fallow areas (Table 1).

Based on the spatial analysis of the 10 products, we derived a consolidated, resampled cropland mask at 250-m resolutions for entire Africa. Since it captured consolidated studies performed by various researchers (Table 1, Fig. 3), covering nominal years 2000 through 2014, it not only captures all the croplands of Africa for nominal year 2014, but also significant portions of non-croplands because a number of datasets in Table 1 are for land use/land cover (LULC) where cropland is a class but has significant non-cropland mix. Working within such a mask (Fig. 3) will help us study all cropland dynamics and their characteristics year after year or season after season for the past MODIS era years (e.g., 2000–2013) as well as for the current study year (2014). However, it raises a question on what if the croplands expand beyond this mask in future years likely to happen. In future studies (2015 and beyond), we need to do a quick study of the areas outside the mask to ascertain any expansion and capture this expansion.

3.3. Classification system and signatures

We mapped four different cropland layers in our product: (1) Cropland extent/area; (2) Irrigated *versus* rainfed; (3) Cropping intensities: single, double, triple, and continuous cropping; (4) Croplands *versus* Cropland Fallows; and crop types (Table 2). There are many differences and inconsistencies in definitions of various global products which is one of the major causes of error distribution (Congalton et al., 2014). The FAO cropland database, for example, defines arable land as land that is under temporary crops (double-cropped areas are counted only once), temporary meadows for mowing or pasture, land under market and kitchen gardens, and land temporarily fallow (less than five years) (Kummu et al., 2012). In the definition, cropland includes all cultivated land under permanent crops, including harvested cropland, crop failure, temporarily fallow or idle land, and cropland used temporarily for pasture; irrigated crop includes all croplands where water from the artificial application is delivered to crops one or more times during crop growing season. Harvest must occur at least once per year (except for plantation crops like tea, coffee, rubber, many varieties of nuts and fruits); rainfed crop includes all croplands where no water from any storage or delivery mechanism is utilized, but crops are not flooded. Cropland fallows are mapped separately.

It is widely accepted that cropland classification accuracies increase when the large areas like continents are stratified and studied separately. After masking out the non-cropland area using 250 m crop mask, the input dataset was subsetted based on the 8 FAO agro-ecological zones (AEZs, Fig. 1). The area in the same AEZ zone has similar characteristics related to land suitability, potential production, and environmental impact. An Agroecological Zone is a land resource mapping unit, defined regarding climate, landform and soils, and land cover, and having a specific range of potentials and constraints for land use. The essential elements in defining an AEZs are the growing period, temperature regime and soil mapping unit.

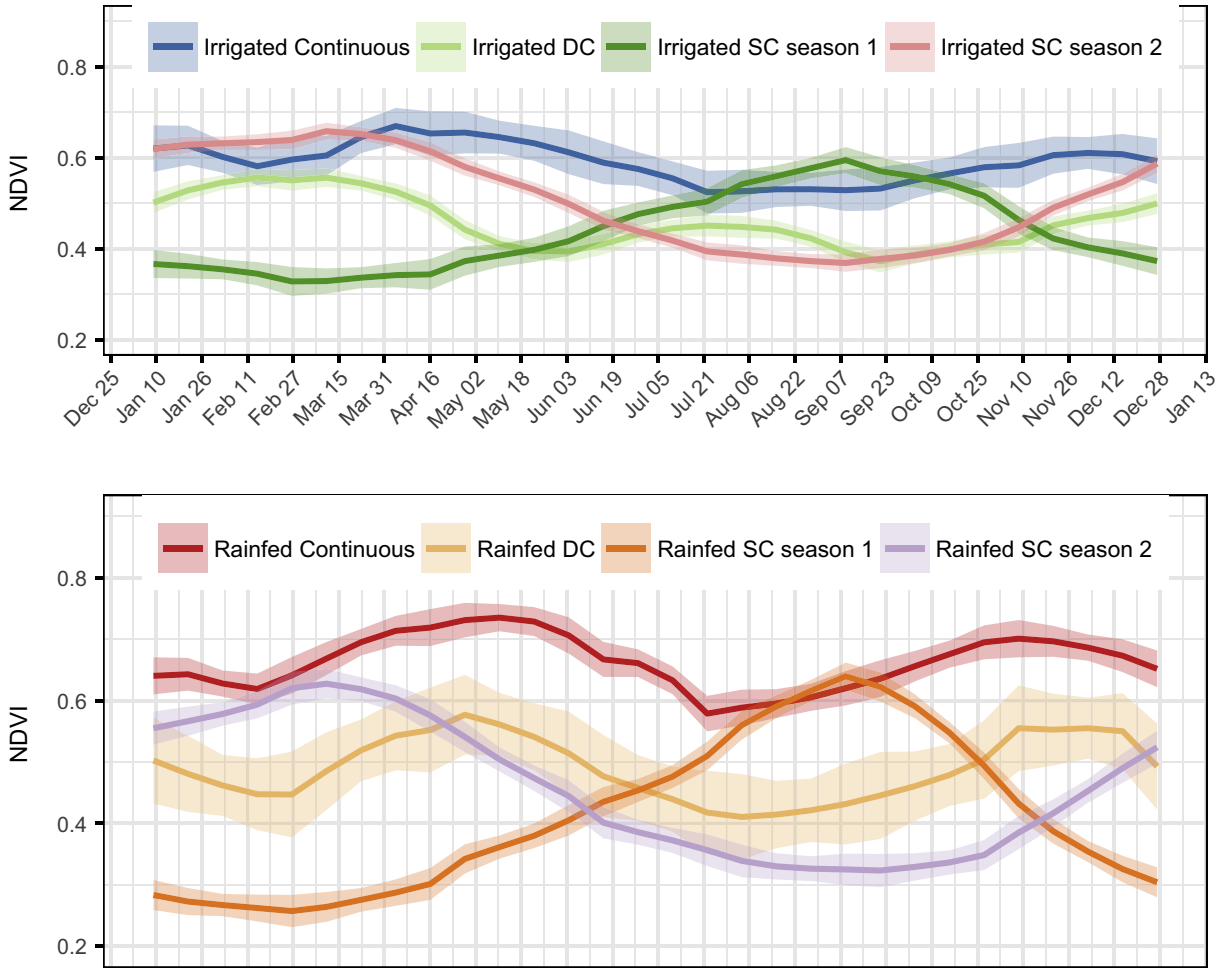

Fig. 4. Ideal spectral signatures of the distinctly separable, unique four irrigated (top) and four rainfed (bottom) classes in agro-ecological zone 3 (AEZ 3), Africa. Illustration on the every 16-day time-series of MODIS 250 m NDVI profiles based on ground data sample knowledge base collected throughout Africa for the year 2014.

Table 2
Description of crops mapped in global cropland product for Africa @ 250-m (GFSAD250).

#	Label	Dominant crop types included	Number of samples
1	Irrigated, SC, season 2	Wheat, barley	28
2	Irrigated, SC, season 1	Maize, rice, millet	14
3	Irrigated, DC,	Rice/chili-vegetable, rice-rice	58
4	Irrigated, Continuous	Sugarcane, plantation	20
5	Rainfed, SC, season 2	Millet, barley, maize, beans, cassava, yam	570
6	Rainfed, SC, season 1	Maize, sorghum, tef, wheat, barley, cassava, yam	257
7	Rainfed, DC,	Rice-rice, maize-maize, rice-beans/ potato/chickpea/pulses	58
8	Rainfed, Continuous	Sugarcane, plantation	57
9	Fallow-lands		10

Note: Season 1: Oct-Mar, season 2: May-Sep.
Only dominant crops are mentioned, since always more than one crop in a single MODIS 250 m pixel (~6.25 ha).

For each agro-ecological zones (AEZs), ideal time-series signatures of unique and distinct classes were established for the irrigated areas and rainfed areas. Our focus was to develop such ideal time-series signatures for classes that are separable from one another. For example, 4 such classes for irrigated and 4 for rainfed were defined in the AEZ 3 (Fig. 4). Indeed, these four classes stood out across AEZs. Classes other than these were either not very distinct/unique, or did not have significant areas and hence were merged into one of the 8 classes. The fallow cropland class was the ninth class, that was common in all AEZs. Establishing the 9 distinct classes (8 classes in Fig. 4 and the ninth class of fallow croplands) allows automated ACMA algorithm coding which in turn will facilitate replicating cropland characteristics year after year or season after season.

The season division is based on cropland calendar and precipitation pattern from ground experience as well as literature (Hentze et al., 2016; Kidane et al., 2012; Kruger, 2006; Lambert et al., 2016; Motha et al., 1980; Waldner et al., 2016), specific in Africa. Some countries, like Zambia, their seasons fall into three periods: Rainy season (December–April), Cool dry season (May–August), Hot dry season (Sept–November). Such case usually affected by a highly unpredictable weather patterns. However, seasonality can be easily discerned using the time-series NDVI (e.g., Fig. 4).

3.4. Creation of reference cropland layer (RCL)

To drive clustering to the massive dataset on terabyte level, all the MODIS tiles covering Africa in 2014 were organized as a large

ImageCollection in Google Earth Engine and then exported in parallel netCDF format (PnetCDF, Li et al. (2003)) on NASA Earth Exchange (NEX, Nemani et al. (2011)) supercomputing platform. Message Passing Interface (MPI) k-means (Zhang et al., 2011) algorithm was applied to do the clustering with 2000 CPUs on NASA AMES super computer. For the total 8 AEZs, K-means cluster results in a total 200 unique clusters for the continental Africa, based on their NDVI time-series profile signal.

A reference cropland layer (RCL) was produced based on zonal classes signature knowledge for the year 2014 (RCL2014), certain class is matched with ideal time series signature library using quantitative spectral matching techniques (QSMTs, Thenkabail et al. (2007)), and is given a preliminary label such as, for example: "rainfed, single, season 1" (Fig. 5). The process is iterated leading to identification and labeling of all 200 classes from the 8 AEZs. The accuracies of the RCL2014 products were based on validation dataset described at Section 2.3.

3.5. Generalization of RCL to ACMA rules

The proposed method uses RCL2014 for developing recursive decision-tree automated cropland mapping algorithm (ACMA) since it consists of the best possible cropland information available for each AEZ of Africa. The construction of decision-tree ACMA algorithm is a procedure that recursively partitions a dataset into smaller subdivisions by a set of tests defined at each branch or node in the tree. The tree is composed of a root node (formed from training data), a set of internal nodes (splits), and a set of terminal nodes (leaves). A zonal tree rules are constructed by recursively partitioning the time series distribution of the reference cropland layer using WEKA (Sharma et al., 2013) and then expert-timed till the derived ACMA generated cropland layer for the year 2014 (ACL2014) accurately matches with RCL2014. In zones where land cover features were misclassified and classification output was considered unsatisfactory, we added training data, redeveloped the decision tree models, and reapplied models.

We used decision tree approach to hierarchically classify crop types. The decision tree for each AEZ consisted of three steps: (a) separately using irrigated/rainfed masks, (b) fallow cropland identification, (c) decision-tree for the primary classes in the individual AEZs. Fallow croplands were filtered out separately for irrigation and rainfed: for irrigation area, area whose NDVI value lower than 0.2 in six months of one calendar year being mapped as cropland fallows; for rainfed area, pixels whose NDVI falls below a threshold during the peak growing seasons of the crop will be coded as cropland fallows.

An example to distinguish MODIS NDVI time series distribution of these classes in Africa are shown in Fig. 4 for AEZ3. Similar eight classes were established across all AEZs. Apart from the eight distinct classes (Fig. 4) across AEZs, a cropland fallow class is also coded based on NDVI falling below a threshold during the critical growing period. The nine classes (Table 2) from the irrigated and rainfed masks of the 8 AEZs are analyzed (e.g., Fig. 5) leading to RCL2014. The knowledge captured in the 9 RCL2014 classes are then coded in ACMA (Fig. 6) to derive ACMA developed cropland layer for the year 2014 (ACL2014). The process of developing the ACMA go through numerous iterations, as illustrated partially in Fig. 6. It involves writing a bunch of simple rules to capture RCL2014 knowledge in the codes and replicate it accurately. Every ACMA rule captures certain percentage of total cropland area and

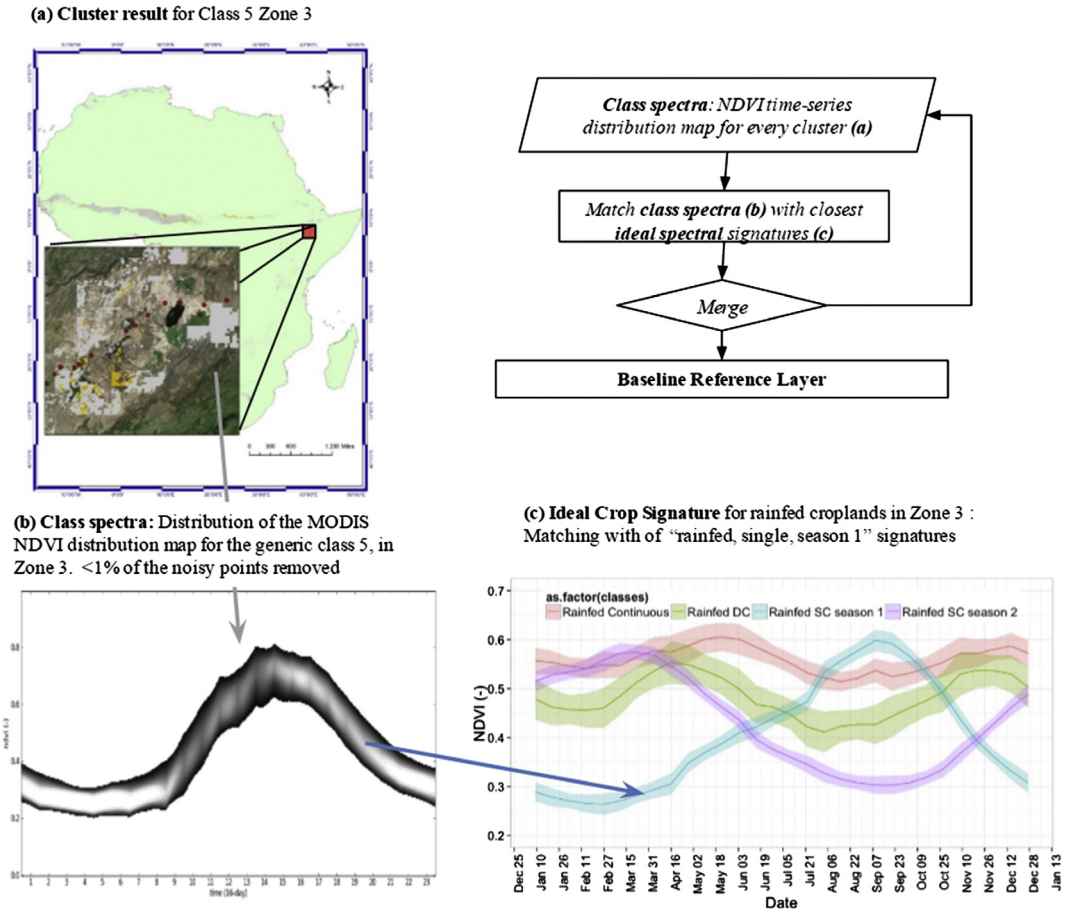

Fig. 5. Quantitative spectral matching (QSM) of a generic class with an ideal spectral signatures.

its characteristics (e.g., irrigated *versus* rainfed or intensity) in each of the nine classes (Table 2) of RCL2014. The process is repeated with numerous additional rules to capture as much cropland area/extent and as many cropland characteristics as possible. If the rule captures non-croplands then the iteration is repeated by tweaking the rule till we can precisely (or near precisely) capture croplands, distinguish them from non-croplands, as well as differentiate irrigated croplands from rainfed croplands or cropping intensities. The process requires several runs to slightly adjust and re-adjust the thresholds till ACL2014 achieves as close a match as possible with RCL2014. The ACMA rules are shown in Fig. 6.

3.6. Ensemble and deployment algorithm on Google Earth Engine for year-to-year-classification

ACMA is a group of decision-trees like what we show in Fig. 6 so we can easily deploy it on Google Earth Engine and run fast for the independent years. Taking MODIS 250-m time-series data as input, we tested ACMA algorithm from 2003 to 2013. This entire ACMA algorithm is made available here: http://geography.wr.usgs.gov/science/croplands/algorithms/africa_250m.html.

The strength of the ACMA algorithm lays in its ability to reproduce cropland products accurately and automatically for the independent years: the past, present, and future. As a result, we used MODIS 250-m time-series data from the year 2003 through 2013 and tested the ACMA algorithm.

3.7. Areal statistics

Full pixel areas (FPAs) are not actual areas. The actual areas are equivalent to sub-pixel areas (SPAs) and are calculated by multiplying SPAs with cropland area fractions (CAFs). This is because a MODIS pixel even when cropped may have a different proportion of crop within the pixel. Thereby:

$$\text{SPAs} = \text{FPAs} \times \text{CAF}$$

where CAFs are determined by taking an average MODIS NDVI image during the growing season and plotting all pixels of the class for this period from the MODIS NDVI image in a brightness-greenness-wetness space (Thenkabailc et al., 2007). The same methodology is adopted here. Also, to get actual areas, one need to re-project MODIS cropland products to appropriate projection. Further, areas are established during different seasons by accounting intensity (single, double, triple, or continuous cropping). Areas cropped twice have areas counted two times a year. Single and continuous have areas computed one time a year.

4. Results

The results start with a reference cropland layer for the year 2014 (RCL2014), followed by the ACMA generated cropland layer for the year 2014 (ACL2014). This will be followed by cropland layers for the 11 independent years 2003–2013 (ACL2003 to ACL2014). Throughout the product validation, area calculations, and comparison with statistical data are presented and discussed.

4.1. Reference cropland layer of Africa for the year 2014 (RCL2014)

4.1.1. Croplands versus non-croplands

Assessment is a key component of map production, especially when remote sensing data are utilized. High-quality reference data

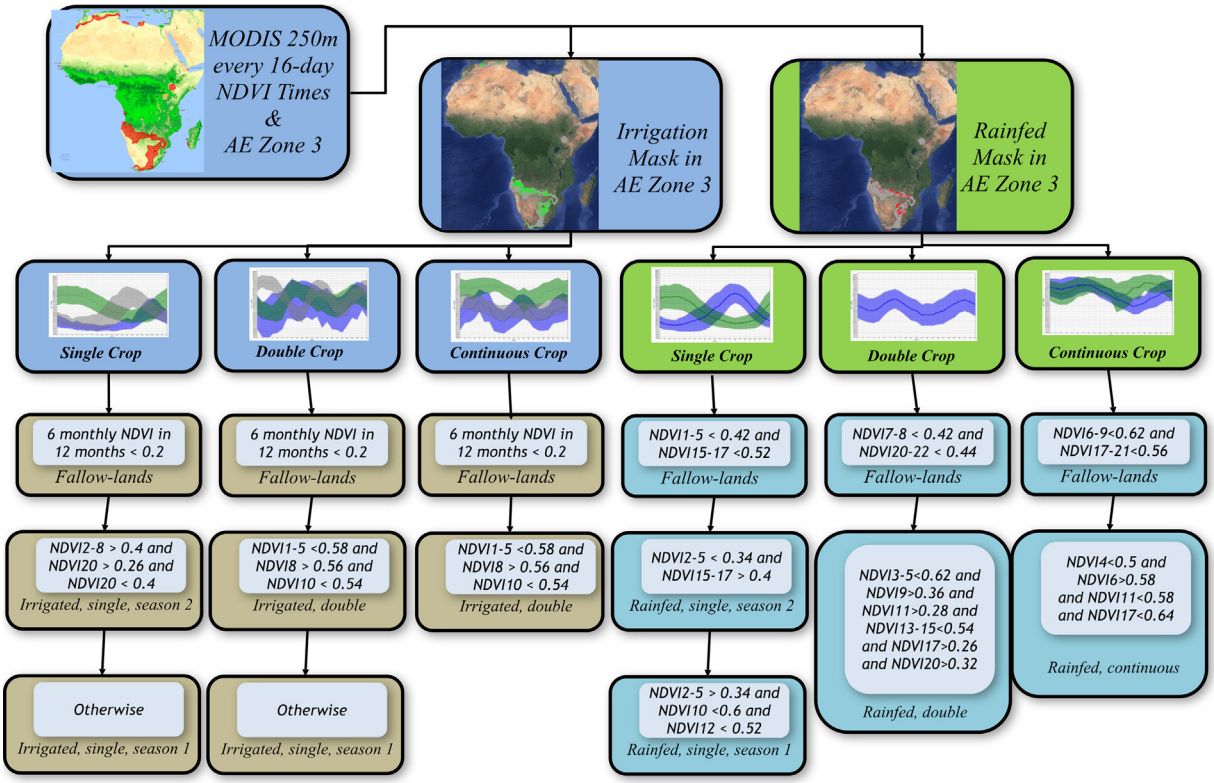

Fig. 6. Example of ACMA algorithm established for AEZ 3. An illustration of the automated cropland mapping algorithm (ACMA) coded and development for the irrigated and rainfed are written so as to capture the knowledge in RCL2014. The process leads to ACMA generated cropland layer for the year 2014 (RCL2014) replicating ACMA generated cropland layer for the year 2014 (ACL2014). ACMA is then applied for other independent years and validated.

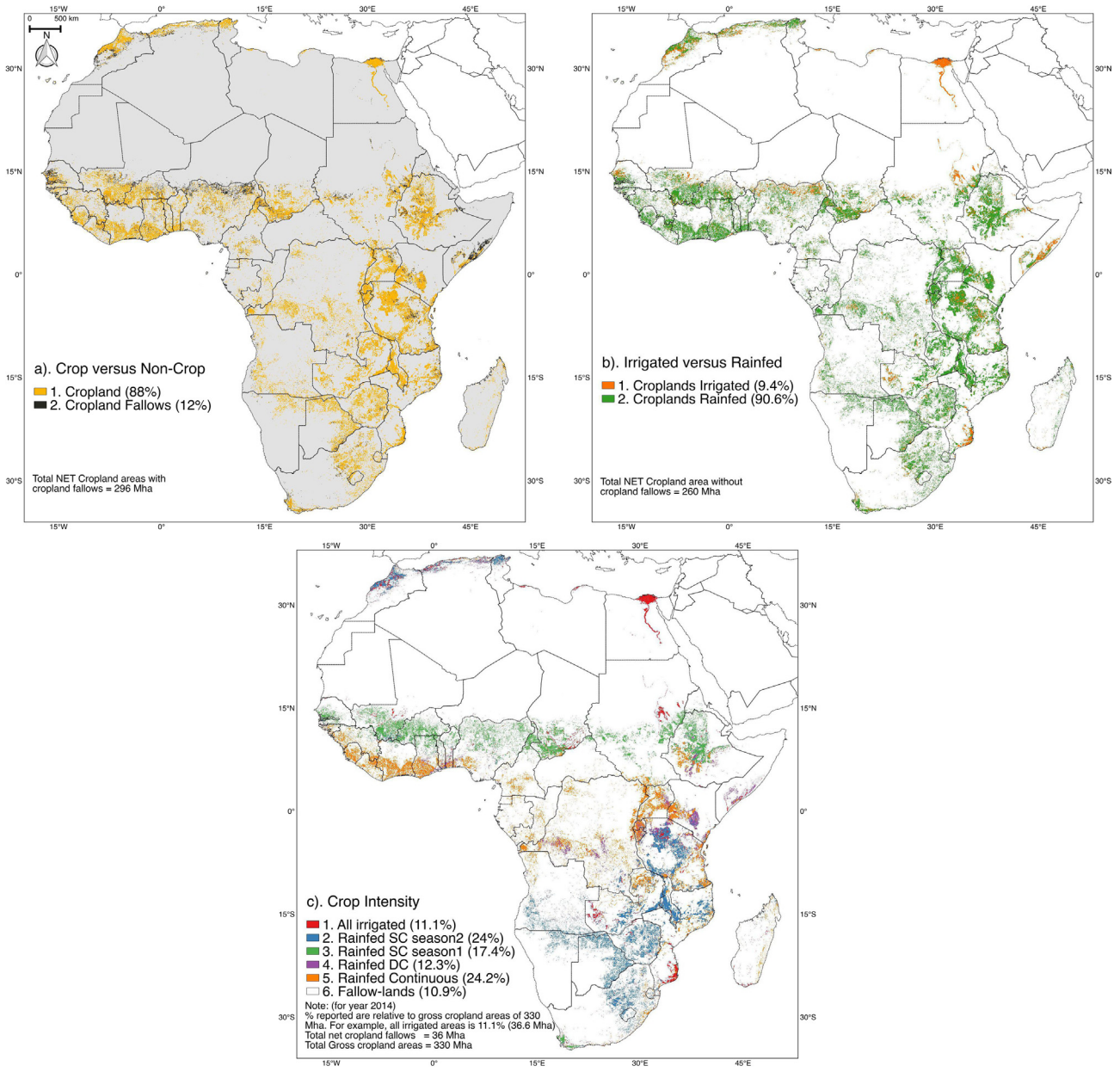

Fig. 7. RCL2014 spatial distribution. Reference cropland layers of Africa for the year 2014 (RCL2014). Three RCL2014 products: (a) Cropland *versus* non-Cropland Layer, (b) Irrigated *versus* Rainfed Layer, and (c) Crop intensities Layer.

is required to be collected at appropriate spatial and temporal scales to perform independent validation (Congalton and Green, 2009). The accuracies of croplands *versus* non-croplands were evaluated for each of the 8 agro-ecological zones (AEZs) and the overall accuracies (OAs) varied between 89% and 100% (Table 3). The accuracy of the resulting cropland products was validated with the global food security support analysis data (GFSAD) project Validation Dataset https://croplands.org/app/data/search, which is a consistent global cropland validation dataset designed for validating cropland products and includes multiple datasets that are ground-based, VHRI based, or sourced from other local detailed studies. In this research a total of 3265 samples, distributed over various agro-ecological zones (AEZs) of Africa, were collected through the crowdsourcing land cover validation tool called cropland.org. Also, the proposed product is compared with cropland statistics derived from other gridded and survey-based data sources. The AEZs 4, 5, 6, 7, and 3, together where about 95% of Africa's croplands exist, have an accuracy of 89–96%. The very high (98–100%) percent accuracies were for AEZs with very low (0.35–4.64%) cropland areas.

The overall agreement of croplands *versus* non-croplands mapped by the 250-m global cropland product of Africa (this study) or GFSAD250 when compared with the gridded dataset from other sources (Table 4), showed that there is an uncertainty between 15% and 25%. Given that all these products are produces using different data, time periods, methods, and approaches, the uncertainties are reasonable.

Besides, a country by country cropland areas was then computed and compared with MIRCA2000 ((Portmann et al., 2010), Fig. 8; The most updated statistics were obtained through personal communication with Portmann and Siebert in 2014 to coincide with our 2014 synthesis). The variability was maximum smaller island nations (e.g., Comoros, Mauritius). Few other countries (e.g., Sierra Leone, Cote dVoire, Chad, Guinea, and Cameroun) also

Table 3
RCL2014 overall accuracies for croplands *versus* non-croplands (product 1). Overall accuracies of RCL2014 product 1 (croplands *versus* non-croplands) based on ground data for Africa in Each AEZs. Overall accuracies (OAs) of the reference cropland layer for the year 2014 (RCL2014) for Africa in each of the 8 agro-ecological zones (AEZs) for croplands *versus* non-croplands (product 1) produced based on MODIS 250 m every 16 day NDVI data, ground data, and spectral matching techniques.

AEZ	Cropland area (Mha)	% of total cropland (%)	Non-cropland area (Mha)	% of total non-crop area (%)	Crop samples (–)	Non-crop samples (–)	Overall accuracy (%)
1	1.0	0.3	134.3	4.8	1	49	100
2	13.8	4.6	1,090.0	39.3	2	48	100
3	24.6	8.3	318.2	11.5	36	209	89
4	106.9	35.9	336.6	12.1	29	204	87
5	94.4	31.7	421.8	15.2	20	208	91
6	28.5	9.6	259.2	9.3	13	279	95
7	26.6	9.0	164.6	5.9	6	235	96
8	1.7	0.6	52.2	1.9	7	243	98
Total	297		2777		114	1378	94

showed significant variability. *R*-square of 0.42 is calculated based on all 55 African Nations (recognized by the United Nations and African Union) (Fig. 9). If we 8 outlier countries, where uncertainty is maximum, the comparisons between: GFSAD250 with MIRCA2000 for the rest 47 countries increases to an *R*-square of 0.69. The GFSAD250 These results clearly imply the ability of GCEA250 to compute cropland areas of Africa and provide country level statistics.

The remote sensing estimates of this work over-estimates areas relative to MIRCA2000 Comoros and Zimbabwe whereas we under-estimate in Cameroon, Côte d'Ivoire and DR Congo. There are reasons for the discrepancy between our remotely sensed products and survey-based statistics like MIRCA2000:

1. Uncertainty in the calculation of MIRCA2000 areas. MIRCA2000 is a derived gridded dataset based on FAOSTAT database (Portmann et al., 2010). FAO compiles the statistics reported by individual countries, which are based on national censuses, agricultural samples, and questionnaire-based surveys with major agricultural producers, independent evaluations (FAO, 2013 The World Bank, 2010). Since each country has its own mechanism, differences in data gathering, and resource limitation, the data lacks objectivity in many countries resulting in data quality issues, particularly in Africa. For example, in 2008/09 in Malawi, cropland extent was estimated by combining household surveys with field measurements derived from a "pacing method" in which the size of crop fields is determined by the number of steps required to walk around them (Dorward and Chirwa, 2010).
2. Application of the ACMA over certain regions have to face the limitation of spatial resolution of MODIS pixels. A typical case is Madagascar, in its slash-and-burn agriculture for pluvial rice which is a predominant component in of cultivation. These fields are easily mixed with neighboring vegetation because lack of cropland management (Messerli and Messerli, 2009), resulting in fallow re-growth in rice fields.

4.1.2. Irrigated versus rainfed croplands

Of the 260 Mha croplands during 2014, 90.6% (236 Mha) was rainfed and just 9.4% (24.5 Mha) was irrigated (Fig. 7b). Africa has 15% of the world population, but just 6% of global irrigated area of 400 Mha (Thenkabail et al., 2009, Thenkabail and Wu, 2012) is in Africa. An overwhelming proportion of the irrigated areas were along the Nile, specifically in Egypt, North Africa, South Africa, along Niger in Mali, and scattered irrigated areas in Southern Africa especially Lake Victoria and Lake Malawi (Fig. 7b). Irrigated *versus* rainfed classification accuracies were evaluated in each of the 8 AEZs and the overall accuracies were between 89% and 94% (Table 5). The accuracy of the irrigated *versus* rainfed cropland products was validated with the global food security support analysis data (GFSAD) project Validation Dataset as discussed before. A country by country comparison of the irrigated areas and the rainfed areas computed by this study with MIRCA2000 reported statistics are plotted in Fig. 10. The *R*-square values were 0.6 for irrigated areas and 0.31 for rainfed areas. Irrigated areas can be computed with great certainty and uncertainties were greatest for the small island nations and few other countries. If the 8 outlier nations are removed, for the rest 47 countries the *R*-square values with MIRCA2000 increased to 0.6 for the rainfed areas. As expected, uncertainties were higher for rainfed croplands and this was mainly as a result of highly fragmented, low biomass croplands that were either confused with grasslands in savannas or regrowth vegetation in the humid tropics. It also has to be noted that MIRCA2000 data which relays on the national statistics also has great degree of uncertainty in rainfed cropland estimates.

4.1.3. Cropping intensity

In both irrigated and rainfed areas of Africa, single crop is overwhelmingly dominant (Fig. 7c, and Table 6). Of the 330 Mha of gross cropland areas during the year 2014, rainfed single crop gross areas was 136.9 Mha (Class 5 and 6 in Table 6). This was followed by rainfed continuous crop gross areas with 79.83 Mha, rainfed double crop gross areas with 40.56 Mha, and irrigated double crop gross areas with 24.17 Mha (Table 6). Gross areas of irrigation single crop during season 1 (4.61 Mha), season 2 (4.84 Mha), and continuous (3.07 Mha) were much smaller. Cropland fallows were 36 Mha during 2014, almost all of that in rainfed croplands with a negligible portion in irrigated croplands (Table 6).

4.1.4. Cropping seasonal layer

Cropland areas are also mapped for two main seasons, continuous crops and a combination of the two seasons (Fig. 8, Table 6). Season 1 (January-May) and season 2 (June-September). Much of the season 1 crops are in Southern Africa and North Africa, while

Table 4
The percent agreement between the global cropland product of Africa @ 250-m (this study) or GFSAD250 when compared with other studies. GlobCover and MODIS MCD12 both have an additional class of mosaic cropland/native vegetation that is added.

GFSAD250AFCE vs. dataset	GRIPC	GLC30	GLC-SHARE	GlobCover (+)	GlobCover (–)	MCD12 (+)	MCD12 (–)
Crop/non-crop agreement %	87.63	86.88	82.95	73.75	72.79	75.88	72.45
Kappa	0.33	0.39	0.48	0.47	0.34	0.46	0.32

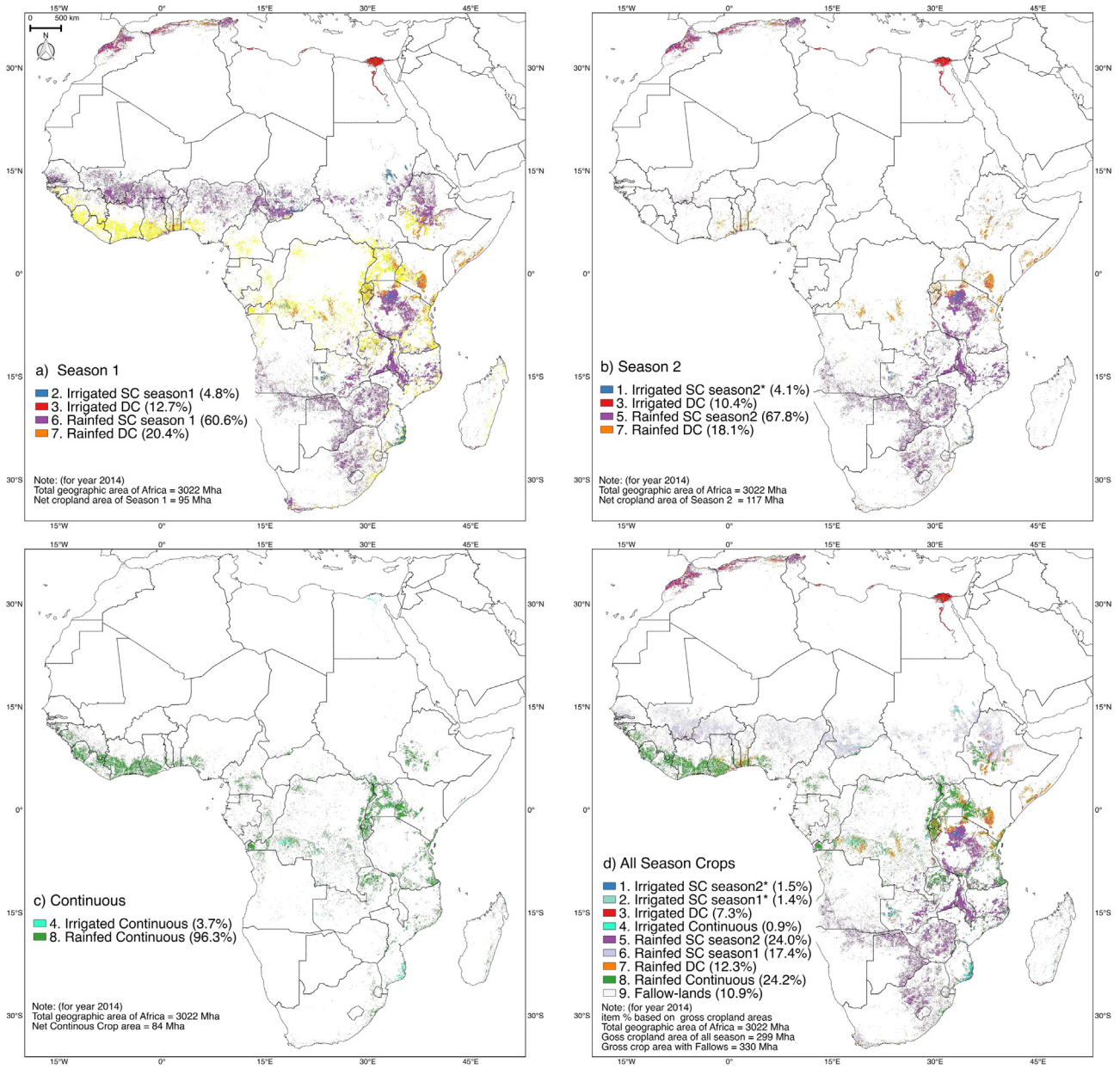

Fig. 8. RCL2014 seasonal cropland layers. Reference cropland product of the year 2014 (RCL2014) for Africa at 250 m generated using MODIS every 16-day time-series data, extensive field knowledge, image classification, and quantitative spectral matching techniques (QSMTs) methods. The top layer shows the croplands from season 1 and season 2 combined, whereas season 1 croplands are shown in bottom left and season 2 croplands are shown in bottom right.

season 2 is mainly distributed in West and Central Africa. Irrigated crops and continuous plantation crops are seen in both seasons, while continuous crops concentrated in West and Central Africa. Overall, for entire Africa, net cropland areas (NCAs) for season 1 was 95 Mha (Fig. 8a, Table 6) and for season 2 was 117 Mha (Fig. 8b, Table 6), and 84 Mha for continuous (Fig. 8c).

4.2. Similarity Matrix comparing ACP2014 with RCL2014

Automated cropland classification algorithm (ACMA) algorithm was applied on MODIS 250 m time-series mega file data cube for the year 2014 (MFDC2014) to obtain an ACMA derived cropland product for the year 2014 (ACL2014) which was then compared with RCL2014, pixel by pixel for entire Africa involving over little over 64.6 million of MODIS 250 m pixels in a similarity matrix (Table 7). The similarity between ACL2014 and RCL2014 was over 90% for every class with overall accuracy of 96% (kappa 0.72). Thus, the ability of ACMA to replicate the nine classes in RCL2014 with high level of accuracies was clearly established.

4.3. ACMA derived annual cropland layers from 2003 to 2014

We applied ACMA algorithm for 11 independent years (2003 through 2014) using MODIS 250 m every 16-day time-series data of these years available on Google Earth Engine. The results as depicted in Fig. 11 showed that the: (a) net cropland areas (NCAs) of Africa increased by about 11 Mha from 2003 to 2014, varying from 253 Mha to 264 Mha; (b) gross cropland areas (GCAs) Africa also increase by about 13 Mha from 2003 to 2014, varying from 323 Mha to 330 Mha; (c) cropland fallows of Africa decreased by about 10 Mha from 2003 to 2014, varying from 43 Mha to 30 Mha. This is, roughly an increase of 1 Mha of croplands per year,

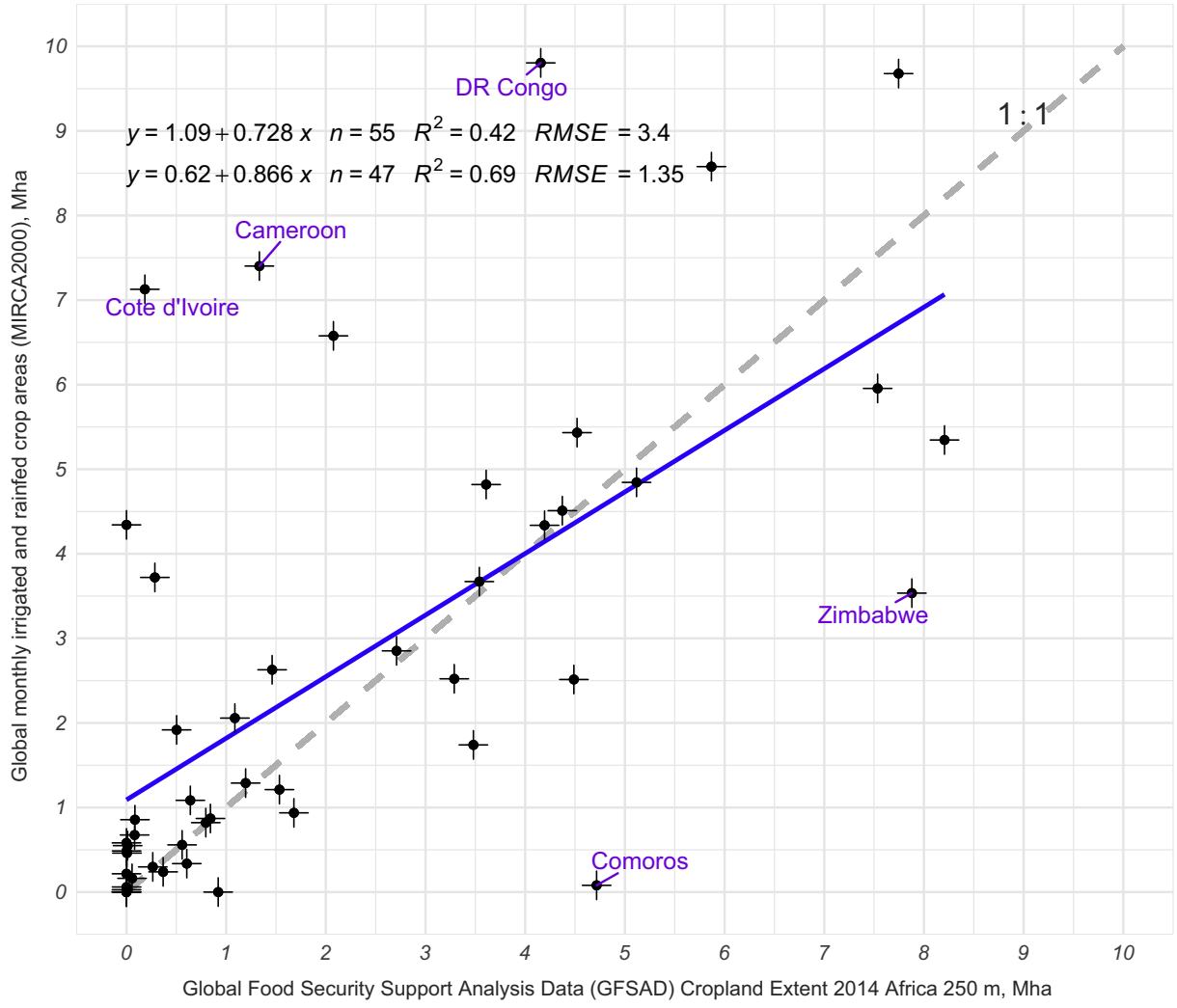

Fig. 9. The global cropland product of Africa @ 250-m (this study) or GFSAD250 derived country by country cropland areas (rainfed + irrigated) of Africa compared with MIRCA2000 (Portmann et al., 2010).

Table 5
RCL2014 overall accuracies for irrigated *versus* rainfed (product 2). Overall accuracies of irrigated *versus* rainfed RCL2014 product 2 (rainfed croplands *versus* irrigated croplands) based on ground data for Africa in Each AEZs. Overall accuracies (OAs) of the reference cropland layer for the year 2014 (RCL2014) for Africa in each of the 8 agro-ecological zones (AEZs) for irrigated *versus* rainfed croplands (product 2) produced based on MODIS 250 m every 16 day NDVI data, ground data, and spectral matching techniques.

AEZ	Irrigated area (Mha)	% of total irrigated area (%)	Rainfed area (Mha)	% of total rainfed area (%)	Irrigated samples (–)	Rainfed samples (–)	Total samples (–)	Overall accuracy (%)
1	0.8	3.62	1.0	0.35	289	11	300	94
2	4.6	20.16	12.7	4.64	117	164	281	89
3	2.4	10.45	22.7	8.26	18	273	291	91
4	6.8	29.39	98.6	35.92	25	267	292	93
5	6.2	26.75	87.1	31.72	20	272	292	86
6	1.7	7.35	26.3	9.57	11	274	285	91
7	0.5	2.25	24.6	8.95	6	280	286	93
8	0.0	0.03	1.6	0.59	2	283	285	92
Total	23		274		488	1824	2312	91

whereas there was a decrease of 1 Mha of cropland fallows per year during the same period. This can only increase further with rapid increase in population and increase food and nutritional demands of the populations. The ability of ACMA algorithm to compute croplands as well as cropland fallows is important one. In drought year cropland fallows increase and cropland areas decrease.

The ability of ACMA to capture variability are depicted between precipitation, ndvi and cropland areas. During drought years, we see dramatic: 1. increase in cropland fallows, and 2. decrease in cropland areas. Also during drought years, there is a significant decrease in the vigor of the existing croplands as illustrated by MODIS 250-m time-series NDVI plots. For example, in the 40,337 hectares' portion of area for AEZ3 depicted in Fig. 13, 57% was crop-

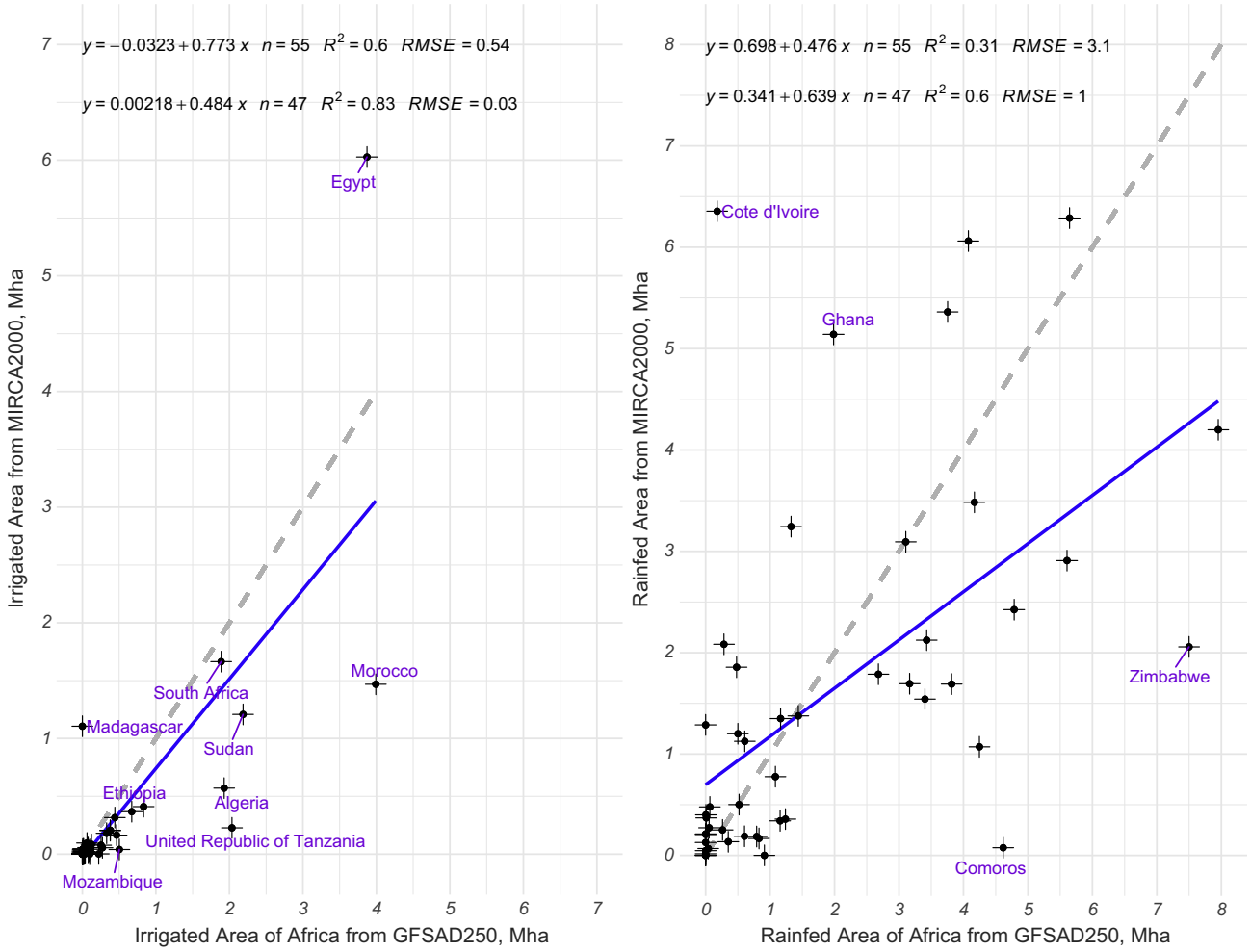

Fig. 10. The global cropland product of Africa @ 250-m (this study) or GFSAD250 reference cropland layer for the year 2014 (RCL2014) irrigation/rainfed country area vs MIRCA2000. (a) Comparison of country-level estimates of cropland area from the new dataset presented in this paper against corresponding data from MIRCA2000 irrigation area. (b) Comparison of country-level estimates of cropland area from the new dataset presented in this paper against corresponding data from MIRCA2000 rainfed area.

Table 6
RCL2014 cropland area seasonal and total statistics of Africa for the year 2014 using MODIS 250 m time-series. The year 2014 cropland area statistics of Africa for the 8 cropland classes and the cropland fallow class. Sub-pixel areas (SPAs) or actual areas were computed for the season 1 (January-May), season 2 (June-September), and for the continuous cropping. Net cropland areas of each season were summed to obtain gross cropland areas from both seasons and for continuous plantation crops.

#	Class	Season 1 (Mha)			Season 2 (Mha)			Continuous (Mha)			Total (Gross Area, Mha)
		FPA	CAF	SPA	FPA	CAF	SPA	FPA	CAF	SPA	SPA
1	Irrigated, SC, season 2				5.09	0.95	4.84				4.84
2	Irrigated, SC, season 1	5.12	0.90	4.61							4.61
3	Irrigated, DC,	13.06	0.92	12.02	13.06	0.93	12.15				24.17
4	Irrigated, Continuous							3.37	0.91	3.07	3.07
5	Rainfed, SC, season 2				93.31	0.85	79.32				79.32
6	Rainfed, SC, season 1	73.82	0.78	57.58							57.58
7	Rainfed, DC,	25.51	0.76	19.39	25.51	0.83	21.17				40.56
8	Rainfed, Continuous							89.70	0.89	79.83	79.83
9	Rainfed, Fallowlands	36.00									
	Net Crop Area (without Fallow)										260
	Net Crop Area (with Fallow)										296
	Gross Crop Area (with Fallow)										330

Note: Season 1: Oct-Mar, season 2: May-Sep.
FPA (Full-Pixel Area) is determined by aggregation of reprojected MODIS Pixels.
CAF (Crop area Fraction) is determined by developing relationship between NDVI of growing season with percent cover.
SPA (Sub-pixel area) is FPA multiplied by CAF.

land fallows during the drought year of 2005, whereas during the normal year of 2008 there was 35% fallow and during the good year of 2006 there were only 4% fallows. Similarly, in 0.62 Mha portion of cropland area in AEZ5 depicted in Fig. 12, 21% was cropland fallows during the drought year of 2005, whereas during the normal year of 2008 there was 12% fallow and during the good year of

Table 7
Similarity matrix between ACMA derived cropland product for the year 2014 (ACL2014) with reference cropland layer (RCL2014).

	Class	RCL2014										Total
		1. Irrigated, SC, season 2	2. Irrigated, SC, season 1	3. Irrigated, DC,	4. Irrigated, Continuous	5. Rainfed, SC, season 2	6. Rainfed, SC, season 1	7. Rainfed, DC,	8. Rainfed, Continuous	9. Fallow-lands	User accuracy (%)	
ACL2014	1. Irrigated, SC, season 2	813,282	21,517								97.4	834,799
	2. Irrigated, SC, season 1	25,992	772,998	10,400							95.5	809,390
	3. Irrigated, DC,		9,455	548,804	13,908						95.9	572,167
	4. Irrigated, Continuous			11,122	514,613						97.9	525,735
	5. Rainfed, SC, season 2					14,207,104	501,841				96.6	14,708,945
	6. Rainfed, SC, season 1					731	12,564,794	511,298			96.1	13,076,823
	7. Rainfed, DC,						145,763	4,570,981	581,352		86.3	5,298,096
	8. Rainfed, Continuous							166,553	11,270,691	208,927	96.8	11,646,171
	9. Fallow-lands								166,553	16,994,878	99.0	17,161,431
	Total	839,274	803,970	570,326	528,521	14,207,835	13,212,398	5,248,832	12,018,596	17,203,805		
	Producer accuracy (%)	96.9	96.1	96.2	97.4	100.0	95.1	87.1	93.8	98.79		
	Overall similarity	0.963										
	Kappa	0.720										

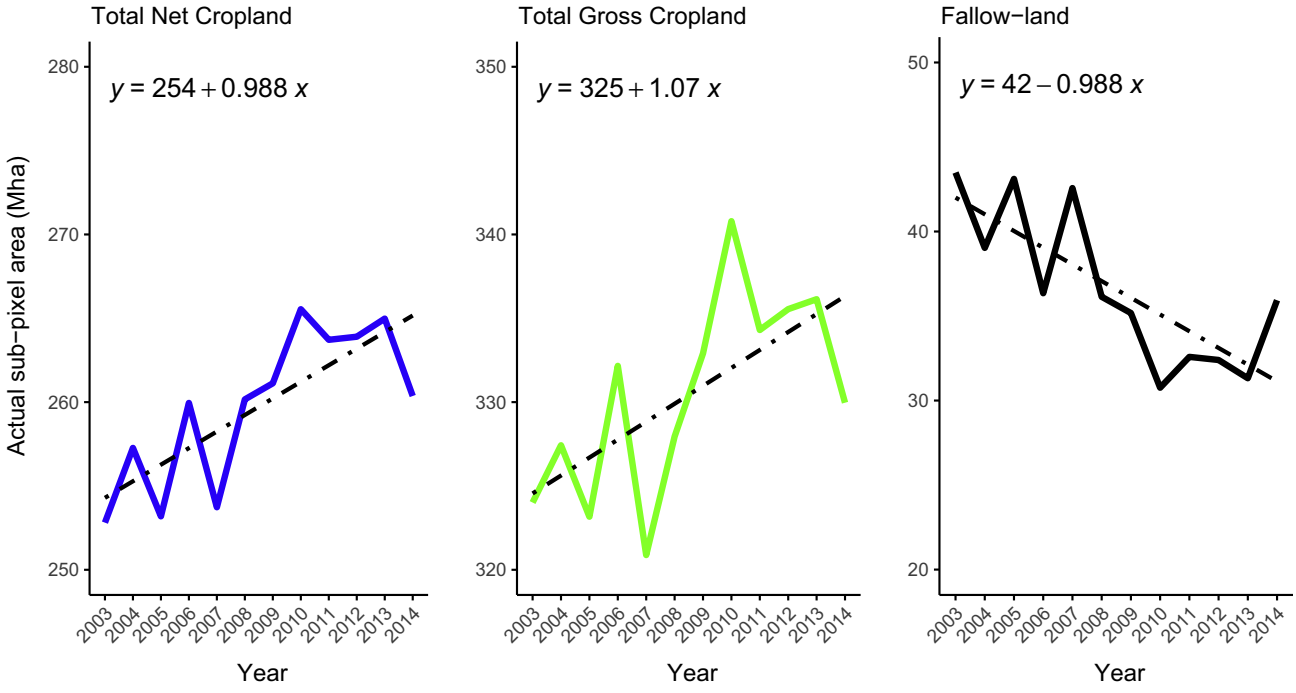

Fig. 11. ACL2003 to ACL2014 derived cropland areas *versus* cropland fallow areas.

2006 there were only 10% fallows. The NDVI vigor trends also clearly depict drought, normal, and good years. Thereby, the ability of ACMA to highlight the combination of the above three factors highlights its value in assessing food security. There cannot be direct accuracy assessment of other years without ground reference data. Nevertheless, We established an online utility called CropRef to generate reference samples using crowdsourcing in our project website https://croplands.org. This allows the use of

Very High Resolution Imagery (VHRI) from DigitalGlobe and similar sub-meter to 5-meter imagery to help generate year specific validation data in the future.

5. Discussion

Efficient annual cropland mapping approaches for operational cropland characterization, mapping, and monitoring must comply with several requirements such as reliability, accuracy, automation and effectiveness. This study demonstrated the ability of the recursive automated cropland mapping algorithm (ACMA) rules to accurately capture the available cropland information over large areas. The process involves the concept of using the knowledge base in the reference cropland layer (RCL) to train and build ACMA algorithm and replicate the RCL accurately and routinely within and across years. Testing and validation of ACMA require us to capture accurate knowledge base from multiple sources (ground samples, photo-interpret, and expert-knowledge). Since the uncertainty of this method depends on the quality and quantity of the reference cropland layer as a primary input, we designed a robust open framework and web-based support system https://croplands.org to support, create, and update this reference cropland layer easily. In this research we trained recursive decision tree ACMA algorithm to achieve very high levels of accuracies (>90%) for 4 irrigated, 4 rainfed, and 1 fallow classes. Larger, and higher quality reference data would facilitate development of accurate automated cropland classification algorithms (ACMA). Here, reference cropland layer (RCL) was used to understand, map and model: (1) knowledge captured from different sources; (2) recursive temporal rules for every pixel; (2) the strengths of the generalized rules.

To achieve greater accuracies, development of ACMA need to be done considering: (a) cropland masks, (b) AEZs, and (c) richness (quality, quantity, and spatial spread) of the reference data. The AEZs help us focus on certain agro-climatic zones and capture their unique characteristics. Along with the AEZ approach, the 250-m Crop mask derived from multiple sources is an important starting point for this study to make significant advance from previous studies. This allowed us to conduct this research by focusing heavily within the cropland mask, where majority (>95%) of the present croplands of Africa exists. Nevertheless, it is important to check any expansion of croplands beyond the existing cropland mask. This requires us to (1) carefully choose the way to stratify the input MODIS data; (2) Collect reference data using crowdsourcing technique and interpret them with correctly. Until we can establish the effective knowledge-based decision-tree and verify the classification output with acceptable accuracies, it is not prudent to apply recursive decision tree ACMA algorithm across the continent with equal certainty. That is all the more reason to approach the ACMA development using AEZs.

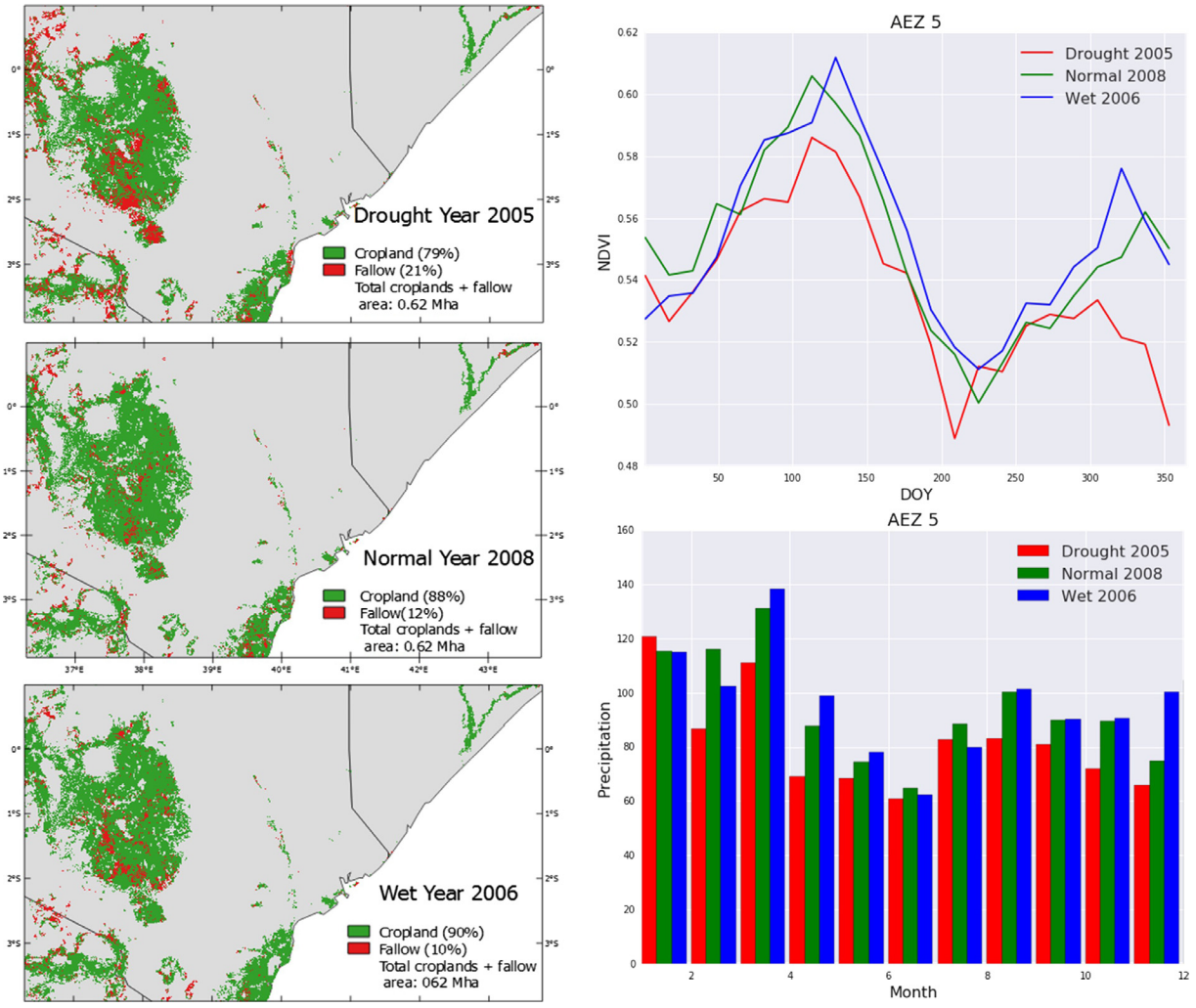

Fig. 12. ACMA derived croplands *versus* cropland fallows for a drought year (2005), normal year (2008), and wet year (2006) in AEZ3. The figure shows spatial distribution of croplands *versus* cropland fallows (left), mean MODIS 250-m NDVI during the three-year (top right) and precipitation (bottom right).

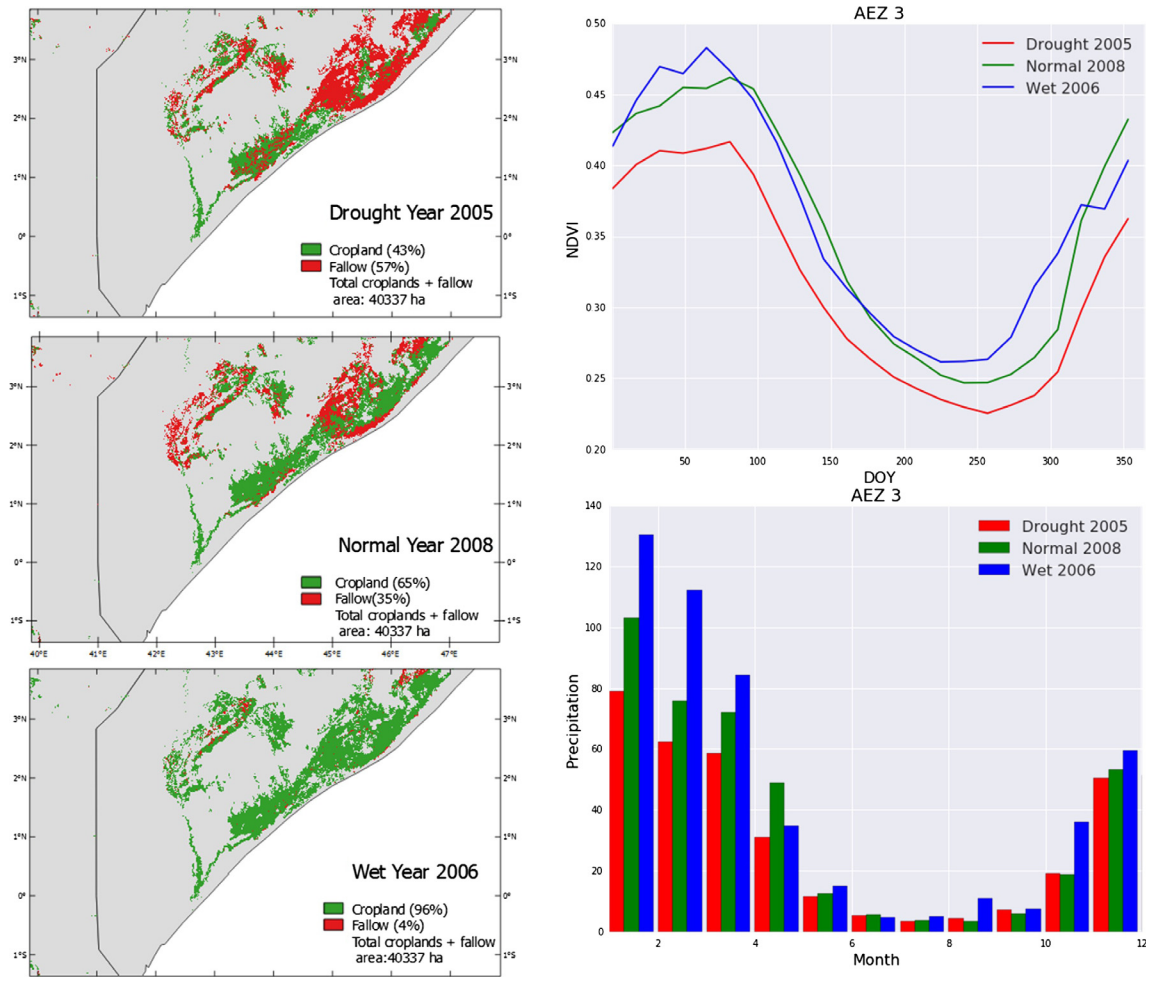

Fig. 13. ACMA derived croplands *versus* cropland fallows for a drought year (2005), normal year (2008), and wet year (2006) in AEZ5. The figure shows spatial distribution of croplands *versus* cropland fallows (left), mean MODIS 250-m NDVI during the three-year (top right) and precipitation (bottom right).

The biggest difficulty in ACMA development and testing is in gathering sufficient training and validation samples to support reliable ACMA coding, rapid product delivery, and accurate product development over such large areas as African continent. A certain class in a particular AEZ may have the lowest producer's and user's accuracies not because of the uncertainty in the classification algorithms but as a result of the poor or biased training and validation datasets. Another challenge comes from up-scaling the local cropland mapping to the continental or global scales through knowledge capture from a zonal decision-tree. Another challenge was to accurately map fallow cropland because of: 1. too few cropland fallow samples, 2. complexity of fallows in defining them, and 3. classification error between cropland fallows and low-density non-cropland vegetation. This might be controlled by better describing the temporal behavior of cropland fallows and updating cropland mask when necessary.

The goal was not to map to many classes where achieving high accuracies becomes complicated, but replicating them year after year (Section 4.3) accurately becomes extremely difficult over very large areas. However, mapping a known number of classes accurately and with ability to replicate year after year also accurately is crucial and meets the important challenge of gathering routine and repetitive cropland statistics over time and space, thus contributing to food security (Figs. 9–11). Often the knowledge of the zonal decision-trees that comes from the reference data sourced from ground samples, photo-interpretation, and expert-knowledge for that zone enriches the recursive ACMA rules for that zone and extrapolation of the same to other zones may not be applicable.

The MODIS 250-m resolution is suitable for national and sub-national applications for the continental level cropland mapping and for deriving cropland statistics at the country and sub-national level. The ability of ACMA to use MODIS time-series data and provide accurate annual updated cropland products is of great interest to the global change science community that benefits from these dynamics because it provides: (a) spatial information content specialized for agriculture; (b) globally consistent and locally relevant information.

In this paper, we present the attempt by tuning classifiers within AEZs rather than entire continent and move to Landsat data in future research. Statistical approaches will have many subjective data gathering techniques that can make the areas collected by this approach uncertain. With different countries having widely varying approaches to statistical data collection, it is hard to standardize. This is the main cause of the scatter we see in Fig. 9. However, for the 47 of the 55 countries the relationship between the remote sensing derived and the MIRCA derived statistics have very good correlation. The advantage of the remote sensing approach is that once the ACMA type algorithms mature, they can be used for routine, repeated, and accurate computation of cropland statistics. In the future, attempts should also be made to better derive statistical areas through standardization and harmonization of data collec-

tion and reporting mechanisms across countries. This along with improved remote sensing products with improved ACMA will better help compare remote sensing derived areas with statistically derived areas.

The use of GEE in data collection to identify reference samples in areas where ground data is lacking does present some new challenges. Sampling and selecting of a homogenous pixel at MODIS scale is not easy (Congalton, 2015), especially working in Africa. This issue can only be controlled through cleaning the input samples with more ancillary data layers where available and remove the outliner as much as possible. Using Google Earth Engine for identifying reference samples to is also debated because they are just interpreted results, not as valuable as the data collected from the field. However, previous research supports the idea that simple, rapid approaches to land cover mapping have benefits. See et al. (2013) found that crowdsourced data from Google Earth delineating the spatial distribution of cropland in Ethiopia had a higher overall accuracy than global land cover datasets. When analyzing the crowdsourced data itself, users underestimate the degree of human impact and there was little difference between experts and non-experts in identifying human impacts.

6. Conclusions

We developed and implemented an automated cropland mapping algorithm (ACMA) using MODIS 250-m 16-day NDVI time-series data. First, a web-based *in situ* reference dataset repository (https://croplands.org/) was developed to collect ground data through field visits, very high spatial resolution data (sub-meter to 5-m), as well as through community by crowdsourcing. Comprehensive knowledge base was then established for Africa using the web repository. Second, a reference cropland layer for the year 2014 (RCL2014) was produced for the entire African continent consisting of 5 crop products: 1. Cropland extent and areas, 2. Irrigated *versus* rainfed croplands, 3. Cropping intensities, 4. Crop type and/or dominance, and 5. Croplands *versus* cropland fallows. Third, decision-tree algorithms were established for the eight individual agro-ecological zones based on RCL2014 knowledge base which was subsequently composed into an automated cropland mapping algorithm (ACMA) applicable for the entire African continent.

The ACMA generated cropland layer for the year 2014 for Africa (ACL2014) when validated showed overall accuracies greater than 89% for each of the eight AEZs. This demonstrated the ability of ACMA to automatically produce cropland products with acceptable accuracies. A country-by-country cropland areas statistics of all 55 African Countries generated from this study was compared with the national census data based MIRCA2000 which were also updated in the year 2014. The relationships showed significant correlations with R-square values between 0.60 and 0.83 for 47 of 55 countries. A pixel-based agreement between the map produced in this study and a number of other studies showed uncertainties varying between 15% and 25%. Overall, for the year 2014, the net cultivated cropland area for the entire African continent was 260 Mha with an additional 36 Mha left fallow. Net cropland area distribution in Africa was 94 Mha during season 1, 117 Mha during season 2, and 84 Mha continuous.

Finally, ACMA algorithm was deployed on the Google Earth Engine cloud computing platform (with executable GEE codes shared at GitHub: https://github.com/suredream/ACM2016), and applied on MODIS data from years 2003 through 2014, to produce annual ACMA generated cropland layers for these years (ACL2003 through ACL2014). The results showed that over 12 years in the African continent there was, on an average, about: (a) 1 Mha/year increase in croplands areas, and (b) 1 Mha/year decrease in cropland fallow areas. The ACMA algorithm clearly demonstrated the ability to accurately capture variations in: A. cropland areas, B. cropland fallow areas, and C. cropland vigor, during drought, normal, and above-normal years routinely and repeatedly year after year over large areas such as for the large continent of Africa. Such ability of the ACMA algorithm clearly provides the needed cropland products for assessing food security. To serve the requirement of resource managers as well as that of the global change research community better, the product and algorithm are made publicly available at: https://croplands.org and http://geography.wr.usgs.gov/science/croplands/algorithms/africa_250m.html.

Acknowledgements

The authors would like to thank following persons for their support: Dr. Felix T. Portman and Dr. Stefan Siebert for providing statistics of MIRCA2000 ((Portmann et al., 2010), also latest statistics through personal communication between Dr. Stefan Siebert and Prasad S. Thenkabail); Dr. Peng Gong for sharing of FROMGLC Validation Dataset (Zhao et al., 2014); Dr. Ryutaro Tateishi for sharing of CEReS Gaia validation data Tateishi et al. (2014), and Dr. Friedl Mark for sharing GRIPC500 dataset for inter-comparison. Special thanks to Dr. Fabio Grita and Dr. Michela Marinelli's help of FAO/CountrySTAT team. Jennifer Dungan (NASA Ames Research Center) and Mutlu Ozdogan (University of Wisconsin-Madison) area acknowledged for reviewing an earlier version of the manuscript. This study was supported by the NASA MEaSUREs (Making Earth System Data Records for Use in Research Environments). The project is funded by NASA MEaSUREs (NNH13AV82I) and the USGS Sales Order number is 29039. We gratefully acknowledge this support. The United States Geological Survey (USGS) provided supplemental funding as well as numerous other direct and indirect support through its Land Change Science (LCS), and Land Remote Sensing (LRS) programs, as well as support from USGS Climate and Land Use Change Mission Area. This research was also supported by Africa Research in Sustainable Intensification for the Next Generation (Africa RISING) program. We would like to thank Dr. Moses Siambi, Research Program Director – Eastern & Southern Africa; Dr. Birhanu Zemadim, Scientist Dr. Ramadjita Tabo, Research Program Director West & Central Africa and Dr. Anthony Whitbread, Research Program Director – Innovation Systems for the Drylands for supporting ground data collection.

References

Arino, O., Gross, D., Ranera, F., Leroy, M., Bicheron, P., Brockman, C., Defourny, P., Vancutsem, C., Achard, F., Durieux, L., Bourg, L., Latham, J., Di Gregorio, A., Witt, R., Herold, M., Sambale, J., Plummer, S., Weber, J.-L., 2007. GlobCover: ESA service for global land cover from MERIS. In: 2007 IEEE International Geoscience and Remote Sensing Symposium. IEEE, pp. 2412–2415.

Bégué, A., Vintrou, E., Saad, A., Hiernaux, P., 2014. Differences between cropland and rangeland MODIS phenology (start-of-season) in Mali. Int. J. Appl. Earth Obser. Geoinform. 31, 167.

Biradar, C.M., Thenkabail, P.S., Noojipady, P., Li, Y., Dheeravath, V., Turral, H., Velpuri, M., Gumma, M.K., Gangalakunta, O.R.P., Cai, X.L., Xiao, X., Schull, M.A., Alankara, R.D., Gunasinghe, S., Mohideen, S., 2009. A global map of rainfed cropland areas (GMRCA) at the end of last millennium using remote sensing. Int. J. Appl. Earth Obser. Geoinform. 11, 114–129.

Chamberlin, J., Jayne, T.S., Headey, D., 2014. Scarcity amidst abundance? Reassessing the potential for cropland expansion in Africa. Food Policy 48, 51–65.

Chen, J., Chen, J., Liao, A., Cao, X., Chen, L., Chen, X., He, C., Han, G., Peng, S., Zhang, W., Lu, M., Tong, X., Mills, J., 2015. Global land cover mapping at 30m resolution: a POK-based operational approach. ISPRS J. Photogramm. Remote Sens. 103, 7.

Clover, J., 2010. Food security in sub-Saharan Africa. Afr. Sec. Stud. 12, 5–15.

Congalton, R., Gu, J., Thenkabail, P., Yadav, K., Ozdogan, M., 2014. Global land cover mapping: a review and uncertainty analysis. Remote Sens. 6, 12070.

Conrad, C., Lamers, J.P.A., Ibragimov, N., LÖw, F., Martius, C., 2016. Analysing irrigated crop rotation patterns in arid Uzbekistan by the means of remote sensing: a case study on post-Soviet agricultural land use. J. Arid Environ. 124, 150.

Congalton, R., Green, K., 2009. Assessing the Accuracy of Remotely Sensed Data: Principles and Practices. CRC/Taylor & Francis, Boca Raton, FL. 183p.

DeFries, R., 2000. Multiple criteria for evaluating machine learning algorithms for land cover classification from satellite data. Remote Sens. Environ. 74, 503.

Congalton, R., 2015. Assessing positional and thematic accuracies of maps generated from remotely sensed data. In: Thenkabail, P. (Ed.), Remote Sensing Handbook, Data Characterization, Classification, and Accuracies, vol. I. CRC/Taylor & Francis, Boca Raton, FL, pp. 583–601.

Defourny, P., Bicheron, P., Brockman, C., Bontemps, S., Van Bogaert, E., Vancutsem, C., Pekel, J.F., Huc, M., Henry, C., Ranera, F., et al., 2009. The first 300 m global land cover map for 2005 using ENVISAT MERIS time series: A product of the GlobCover system. In: Proceedings of the 33rd International Symposium on Remote Sensing of Environment, Stresa, Italy, vol. 48.

Deng, C., Wu, C., 2013. The use of single-date MODIS imagery for estimating large-scale urban impervious surface fraction with spectral mixture analysis and machine learning techniques. ISPRS J. Photogramm. Remote Sens. 86, 100.

Dheeravath, V., Thenkabail, P.S., Thenkabail, P.S., Noojipady, P., Chandrakantha, G., Reddy, G.P.O., Gumma, M.K., Biradar, C.M., Velpuri, M., Gumma, M.K., 2010. Irrigated areas of India derived using MODIS 500 m time series for the years 2001–2003. ISPRS J. Photogramm. Remote Sens. 65, 42.

Ding, Y., Ding, Y., Zhao, K., Zhao, K., Zheng, X., Zheng, X., Jiang, T., Jiang, T., 2014. Temporal dynamics of spatial heterogeneity over cropland quantified by time-series NDVI, near infrared and red reflectance of Landsat 8 OLI imagery. Int. J. Appl. Earth Obser. Geoinform. 30, 139.

Dong, J., Xiao, X., Kou, W., Qin, Y., Zhang, G., Li, L., Jin, C., Zhou, Y., Wang, J., Biradar, C., Liu, J., Moore, B., 2015. Tracking the dynamics of paddy rice planting area in 1986–2010 through time series Landsat images and phenology-based algorithms. Remote Sens. Environ. 160, 99.

Dorward, A., Chirwa, E., 2010. A Review of Methods for Estimating Yield and Production Impacts.

Duro, D.C., Franklin, S.E., Dub, M.G., 2012. A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery. Remote Sens. Environ. 118, 259–272.

Egorov, A.V., Hansen, M.C., Roy, D.P., Kommareddy, A., Potapov, P.V., 2015. Image interpretation-guided supervised classification using nested segmentation. Remote Sens. Environ. 165, 135–147.

Esch, T., Metz, A., Marconcini, M., Keil, M., 2014. Combined use of multi-seasonal high and medium resolution satellite imagery for parcel-related mapping of cropland and grassland. Int. J. Appl. Earth Obser. Geoinform. 28, 230–237.

FAO, I., Fischer, G., Nachtergaele, F., Prieler, S., Velthuizen, H.T.V., Verelst, L., Wiberg, D., 2012. Global Agro-ecological Zones (GAEZ v3.0).

FAO, 2013. Rome. FAOSTAT database. Food and Agriculture Organization of the United Nations, Rome, Italy.

Foody, G.M., 2010. Assessing the accuracy of land cover change with imperfect ground reference data. Remote Sens. Environ. 114, 2271–2285.

Friedl, M.A., Brodley, C.E., 1997. Decision tree classification of land cover from remotely sensed data. Remote Sens. Environ. 61, 399–409.

Friedl, M.A., McIver, D.K., Hodges, J., Zhang, X.Y., 2002. Global land cover mapping from MODIS: algorithms and early results. Remote Sens. Environ., 287–302

Fritz, S., McCallum, I., Schill, C., Perger, C., Grillmayer, R., Achard, F., Kraxner, F., Obersteiner, M., 2009. Geo-Wiki.Org: the use of crowdsourcing to improve global land cover. Remote Sens. 1, 345–354.

Fritz, S., See, L., 2008. Identifying and quantifying uncertainty and spatial disagreement in the comparison of Global Land Cover for different applications. Global Change Biol. 14, 1057–1075.

Fritz, S., See, L., Rembold, F., 2010. Comparison of global and regional land cover maps with statistical information for the agricultural domain in Africa. Int. J. Remote Sens. 31 (9), 2237–2256. http://dx.doi.org/10.1080/01431160902946598.

Fritz, S., See, L., McCallum, I., Schill, C., Obersteiner, M., van der Velde, M., Boettcher, H., Havlík, P., Achard, F., 2011a. Highlighting continued uncertainty in global land cover maps for the user community. Environ. Res. Lett. 6, 044005.

Fritz, S., You, L., Bun, A., See, L., McCallum, I., Schill, C., Perger, C., Liu, J., Hansen, M., Obersteiner, M., 2011b. Cropland for sub-Saharan Africa: a synergistic approach using five land cover data sets. Geophys. Res. Lett. 38.

Funk, C.C., Peterson, P.J., Landsfeld, M.F., Pedreros, D.H., Verdin, J.P., Rowland, J.D., Romero, B.E., Husak, G.J., Michaelsen, J.C., Verdin, A.P., 2014. A Quasi-global Precipitation Time Series for Drought Monitoring. US Geological Survey Data Series 832.

Gerland, P., Raftery, A.E., Ev Ikova, H., Li, N., Gu, D., Spoorenberg, T., Alkema, L., Fosdick, B.K., Chunn, J., Lalic, N., Bay, G., Buettner, T., Heilig, G.K., Wilmoth, J., 2014. World population stabilization unlikely this century. Science 346, 234–237.

Giri, C., Zhu, Z., Reed, B., 2005. A comparative analysis of the Global Land Cover 2000 and MODIS land cover data sets. Remote Sens. Environ. 94, 123–132.

Gong, P., Wang, J., Yu, L., Zhao, Y., Zhao, Y., Liang, L., Niu, Z., Huang, X., Fu, H., Liu, S., Li, C., Li, X., Fu, W., Liu, C., Xu, Y., Wang, X., Cheng, Q., Hu, L., Yao, W., Zhang, H., Zhu, P., Zhao, Z., Zhang, H., Zheng, Y., Ji, L., Zhang, Y., Chen, H., Yan, A., Guo, J., Yu, L., Wang, L., Liu, X., Shi, T., Zhu, M., Chen, Y., Yang, G., Tang, P., Xu, B., Giri, C., Clinton, N., Zhu, Z., Chen, J., Chen, J., 2013. Finer resolution observation and monitoring of global land cover: first mapping results with Landsat TM and ETM + data. Int. J. Remote Sens. 34 (7), 2607–2654. http://dx.doi.org/10.1080/01431161.2012.748992.

Gumma, M.K., Gumma, M.K., Thenkabail, P.S., Maunahan, A., Islam, S., Nelson, A., 2014. Mapping seasonal rice cropland extent and area in the high cropping intensity environment of Bangladesh using MODIS 500m data for the year 2010. ISPRS J. Photogramm. Remote Sens. 91, 98.

Haack, B., Mahabir, R., Kerkering, J., 2014. Remote sensing-derived national land cover land use maps: a comparison for Malawi. Geocarto Int. 30, 270–292.

Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S.V., Goetz, S.J., Loveland, T.R., Kommareddy, A., Egorov, A., Chini, L., Justice, C.O., Townshend, J.R.G., 2013. High-resolution global maps of 21st-century forest cover change. Science 342, 850–853.

Hansen, M.C., Reed, B., 2010. A comparison of the IGBP DISCover and University of Maryland 1 km global land cover products. Int. J. Remote Sens. 21, 1365–1373.

Helmholz, P., Rottensteiner, F., Heipke, C., 2014. Semi-automatic verification of cropland and grassland using very high resolution mono-temporal satellite images. ISPRS J. Photogramm. Remote Sens. 97, 204–218.

Hentze, K., Thonfeld, F., Menz, G., 2016. Evaluating crop area mapping from MODIS time-series as an assessment tool for Zimbabwes fast track land reform programme. PLoS ONE 11, e0156630.

Herold, M., Mayaux, P., Woodcock, C.E., Baccini, A., Schmullius, C., 2008. Some challenges in global land cover mapping: an assessment of agreement and accuracy in existing 1 km datasets. Remote Sens. Environ. 112, 2538–2556.

Herold, M., Woodcock, C.E., Di Gregorio, A., Mayaux, P., Belward, A.S., Latham, J., Schmullius, C.C., 2006. A joint initiative for harmonization and validation of land cover datasets. IEEE Trans. Geosci. Remote Sens. 44, 1719–1727.

Jamali, S., Seaquist, J., Eklundh, L., Ardö, J., 2014. Automated mapping of vegetation trends with polynomials using NDVI imagery over the Sahel. Remote Sens. Environ. 141, 79.

Jayne, T.S., Chamberlin, J., Headey, D.D., 2014. Land pressures, the evolution of farming systems, and development strategies in Africa: a synthesis. Food Policy 48, 1–17.

Jeganathan, C., Dash, J., Atkinson, P.M., 2014. Remotely sensed trends in the phenology of northern high latitude terrestrial vegetation, controlling for land cover change and vegetation type. Remote Sens. Environ. 143, 154.

Kalensky, Z.D., 1998. AFRICOVER land cover database and map of Africa. Can. J. Remote Sens. 24 (3), 292–297. http://dx.doi.org/10.1080/07038992.1998.10855250.

Kidane, Y., Stahlmann, R., Beierkuhnlein, C., 2012. Vegetation dynamics, and land use and land cover change in the Bale Mountains, Ethiopia. Environ. Monit. Assess. 184, 7473–7489.

Kruger, A.C., 2006. Observed trends in daily precipitation indices in South Africa: 1910–2004. Int. J. Climatol. 26, 2275–2285.

Kummu, M., De Moel, H., Porkka, M., Siebert, S., Varis, O., Ward, P.J., 2012. Lost food, wasted resources: global food supply chain losses and their impacts on freshwater, cropland, and fertiliser use. Sci. Total Environ. 438, 477.

Lambert, M.-J., Waldner, F., Defourny, P., 2016. Cropland mapping over Sahelian and Sudanian agrosystems: a knowledge-based approach using PROBA-V time series at 100-m. Remote Sens. 8, 232.

Lary, D.J., Alavi, A.H., Walker, A.L., Gandomi, A.H., 2016. Machine learning in geosciences and remote sensing. Geosci. Front. 7, 3.

Latham, J., Cumani, R., Rosati, I., Bloise, M., 2014. Global Land Cover Share (GLC-SHARE) Database Beta-release Version 1.0-2014. FAO, Rome.

Leroux, L., Jolivot, A., Bégué, A., Seen, D., Zoungrana, B., 2014. How reliable is the MODIS land cover product for crop mapping Sub-Saharan agricultural landscapes? Remote Sens. 6, 8541–8564.

Li, J., Liao, W.-k., Choudhary, A., Ross, R., Thakur, R., Gropp, W., Latham, R., Siegel, A., Gallagher, B., Zingale, M., 2003. Parallel netCDF: a high-performance scientific I/O interface. In: Supercomputing, 2003 Acm/Ieee Conference. IEEE, pp. 39–39.

Lobell, D.B., Asner, G.P., 2004. Cropland distributions from temporal unmixing of MODIS data. Remote Sens. Environ. 93, 412–422.

Lobell, D.B., Burke, M.B., Tebaldi, C., Mastrandrea, M.D., Falcon, W.P., Naylor, R.L., 2008. Prioritizing climate change adaptation needs for food security in 2030. Science 319, 607.

McCallum, I., Obersteiner, M., Nilsson, S., Shvidenko, A., 2006. A spatial comparison of four satellite derived 1km global land cover datasets. Int. J. Appl. Earth Obser. Geoinform. 8, 246–255.

Messerli, P., Messerli, P., 2009. Use of sensitivity analysis to evaluate key factors for improving slash-and-burn cultivation systems on the eastern escarpment of Madagascar. Mount. Res. Dev. 20, 32–41.

Motha, R.P., Leduc, S.K., Steyaert, L.T., Sakamoto, C.M., Strommen, N.D., Motha, R.P., Leduc, S.K., Steyaert, L.T., Sakamoto, C.M., Strommen, N.D., 1980. Precipitation patterns in West Africa. Mon. Weather Rev. 2 (108), 1567–1578. http://dx.doi.org/10.1175/1520-0493(1980)108<1567:PPIWA>2.0.CO.

Mountrakis, G., Im, J., Ogole, C., 2011. Support vector machines in remote sensing: a review. ISPRS J. Photogramm. Remote Sens. 66, 247.

Müller, H., Rufin, P., Griffiths, P., Barros Siqueira, A.J., Hostert, P., 2015. Mining dense Landsat time series for separating cropland and pasture in a heterogeneous Brazilian savanna landscape. Remote Sens. Environ. 156, 490–499.

Nemani, R., Votava, P., Michaelis, A., Melton, F., Milesi, C., 2011. Collaborative supercomputing for global change science. Eos, Trans. Am. Geophys. Union 92, 109–110.

Ozdogan, M., Gutman, G., 2008. A new methodology to map irrigated areas using multi-temporal MODIS and ancillary data: an application example in the continental US. Remote Sens. Environ. 112, 3520–3537.

Pan, Z., Huang, J., Zhou, Q., Wang, L., Cheng, Y., Zhang, H., Blackburn, G.A., Yan, J., Liu, J., 2015. Mapping crop phenology using NDVI time-series derived from HJ-1 A/B data. Int. J. Appl. Earth Obser. Geoinform. 34, 188–197.
Pantazi, X.E., Moshou, D., Alexandridis, T., Mouazen, A.M., Whetton, R.L., 2016. Wheat yield prediction using machine learning and advanced sensing techniques. Comput. Electron. Agric. 121, 57.
PeñArancibia, J.L., Mainuddin, M., Kirby, J.M., Chiew, F.H.S., McVicar, T.R., Vaze, J., 2016. Assessing irrigated agriculture's surface water and groundwater consumption by combining satellite remote sensing and hydrologic modelling. Sci. Total Environ. 542, 372–382.
Pittman, K., Hansen, M.C., Becker-Reshef, I., Potapov, P.V., Justice, C.O., 2010. Estimating global cropland extent with multi-year MODIS data. Remote Sens. 2, 1844–1863.
Portmann, F.T., Siebert, S., Döll, P., 2010. MIRCA2000—global monthly irrigated and rainfed crop areas around the year 2000: a new high-resolution data set for agricultural and hydrological modeling. Global Biogeochem. Cycles 24, 1–24.
Qiu, B., Zhong, M., Tang, Z., Wang, C., 2014. A new methodology to map double-cropping croplands based on continuous wavelet transform. Int. J. Appl. Earth Obser. Geoinform. 26, 97–104.
Radoux, J., Lamarche, C., Van Bogaert, E., Bontemps, S., Brockmann, C., Defourny, P., 2014. Automated training sample extraction for global land cover mapping. Remote Sens. 6, 3965–3987.
Rembold, F., Carnicelli, S., Nori, M., Ferrari, G.A., 2000. Use of aerial photographs, Landsat TM imagery and multidisciplinary field survey for land-cover change analysis in the lakes region (Ethiopia). Int. J. Appl. Earth Obser. Geoinform. 2, 181–189.
Salmon, J.M., Friedl, M.A., Frolking, S., Wisser, D., Douglas, E.M., 2015. Global rain-fed, irrigated, and paddy croplands: a new high resolution map derived from remote sensing, crop inventories and climate data. Int. J. Appl. Earth Obser. Geoinform. 38, 321–334.
See, L., Comber, A., Salk, C., Fritz, S., van der Velde, M., Perger, C., Schill, C., McCallum, I., Kraxner, F., Obersteiner, M., 2013. Comparing the quality of crowdsourced data contributed by expert and non-experts. PLoS ONE 8, e69958.
Shalaby, A., Tateishi, R., 2007. Remote sensing and GIS for mapping and monitoring land cover and land-use changes in the Northwestern coastal zone of Egypt. Appl. Geogr. 27, 28–41.
Shao, Y., Lunetta, R.S., 2012. Comparison of support vector machine, neural network, and CART algorithms for the land-cover classification using limited training data points. ISPRS J. Photogramm. Remote Sens. 70, 78.
Sharma, R., Ghosh, A., Joshi, P.K., 2013. Decision tree approach for classification of remotely sensed satellite data using open source support. J. Earth Syst. Sci. 122, 1237.
Soulard, C.E., Acevedo, W., Auch, R.F., Sohl, T.L., Drummond, M.A., Sleeter, B.M., Sorenson, D.G., Kambly, S., Wilson, T.S., Taylor, J.L., et al., 2014, Land Cover Trends Dataset, 1973–2000: U.S. Geological Survey Data Series 844, 10 p., https://dx.doi.org/10.3133/ds844.
Tateishi, R., Hoan, N.T., Kobayashi, T., 2014. Production of global land cover DataGLCNMO2008. J. Geogr. Geol. 6, 99.
Tatsumi, K., Yamashiki, Y., Canales Torres, M.A., Taipe, C.L.R., 2015. Crop classification of upland fields using Random forest of time-series Landsat 7 ETM+ data. Comput. Electron. Agric. 115, 171–179.
Teluguntla, P., Ryu, D., George, B., Walker, J.P., Malano, H.M., 2015a. Mapping flooded rice paddies using time series of modis imagery in the Krishna River Basin, India. Remote Sens. 7, 8858–8882.
Teluguntla, P., Thenkabail, P., Xiong, J., Gumma, M.K., Giri, C., Milesi, C., Ozdogan, M., Congalton, R., Yadav, K., 2015b. Global food security support analysis data at nominal 1 km (GFSAD1km) derived from remote sensing in support of food security in the twenty-first century: current achievements and future possibilities. In: Thenkabail, P.S. (Ed.). CRC Press, Boca Raton, London, New York, pp. 131–160 (Chapter 6).
Teluguntla, P., Thenkabail, P.S., Xiong, J., Gumma, M.K., Congalton, R.G., Oliphant, A., Poehnelt, J., Yadav, K., Rao, M.N., Massey, R., 2016. Spectral matching techniques (SMTs) and automated cropland classification algorithms (ACCAs) for mapping croplands of Australia using MODIS 250-m time-series (2000–2015) data. Int. J. Digital Earth, 1–36.
Thenkabail, P., GangadharaRao, P., Biggs, T., 2007. Spectral matching techniques to determine historical land-use/land-cover (LULC) and irrigated areas using time-series 0.1-degree AVHRR Pathfinder datasets. Photogramm. Eng. Remote Sens. 73, 1029–1040.
Thenkabail, P.S., Wu, Z., 2012. An automated cropland classification algorithm (ACCA) for Tajikistan by combining Landsat, MODIS, and secondary data. Remote Sens. 4, 2890–2918.
Thenkabailc, P.S., Biradar, C.M., Noojipady, P., Cai, X., Dheeravath, V., Li, Y., Velpuri, M., Gumma, M., Pandey, S., 2007. Sub-pixel area calculation methods for estimating irrigated areas. Sensors 7, 2519.
Thenkabail, P.S., Biradar, C.M., Noojipady, P., Dheeravath, V., Li, Y., Velpuri, M., Gumma, M., Gangalakunta, O.R.P., Turral, H., Cai, X., et al., 2009. Global irrigated area map (GIAM), derived from remote sensing, for the end of the last millennium. Int. J. Remote Sens. 30 (14).
Tsendbazar, N.E., Tsendbazar, N.E., de Bruin, S., Herold, M., Herold, M., 2015. Assessing global land cover reference datasets for different user communities. ISPRS J. Photogramm. Remote Sens. 103, 93.
Vancutsem, C., Marinho, E., Kayitakire, F., See, L., Fritz, S., 2012. Harmonizing and combining existing land cover/land use datasets for cropland area monitoring at the African continental scale. Remote Sens. 5 (1), 19–41. http://dx.doi.org/10.3390/rs5010019.
Vintrou, E., Desbrosse, A., Bégué, A., Traoré, S., 2012. Crop area mapping in West Africa using landscape stratification of MODIS time series and comparison with existing global land products. Int. J. Appl. Earth Obser. Geoinform. 14, 83–93.
Waldner, F., Canto, G.S., Defourny, P., 2015. Automated annual cropland mapping using knowledge-based temporal features. ISPRS J. Photogramm. Remote Sens. 110, 1–13.
Waldner, F., De Abelleyra, D., Verén, S.R., Zhang, M., Wu, B., Plotnikov, D., Bartalev, S., Lavreniuk, M., Skakun, S., Kussul, N., Le Maire, G., Dupuy, S., Jarvis, I., Defourny, P., 2016. Towards a set of agrosystem-specific cropland mapping methods to address the global cropland diversity. Int. J. Remote Sens. 37, 3196–3231.
Wang, J., Zhao, Y., Yu, L., Li, C., Yu, Liu, D., Gong, P., 2015. Mapping global land cover in 2001 and 2010 with spatial-temporal consistency at 250m resolution. ISPRS J. Photogramm. Remote Sens. 103, 38.
Were, K.O., Dick, Ø.B., Singh, B.R., 2013. Remotely sensing the spatial and temporal land cover changes in Eastern Mau forest reserve and Lake Nakuru drainage basin, Kenya. Appl. Geogr. 41, 75–86.
Wu, Z., Verdin, J.P., Thenkabail, P.S., Verdin, J.P., 2014. Automated cropland classification algorithm (ACCA) for California using multi-sensor remote sensing. Photogramm. Eng. Remote Sens. 80, 81.
Yan, L., Roy, D.P., 2014. Automated crop field extraction from multi-temporal Web Enabled Landsat Data. Remote Sens. Environ. 144, 42.
Zhang, G., Zhang, G., Xiao, X., Dong, J., Kou, W., Kou, W., Jin, C., Jin, C., Qin, Y., Zhou, Y., Qin, Y., Zhou, Y., Wang, J., Menarguez, M.A., Biradar, C., 2015. Mapping paddy rice planting areas through time series analysis of MODIS land surface temperature and vegetation index data. ISPRS J. Photogramm. Remote Sens. 106, 157.
Zhang, J., Wu, G., Hu, X., Li, S., Hao, S., 2011. A Parallel K-Means Clustering Algorithm with MPI. IEEE.
Zhao, Y., Gong, P., Yu, L., Hu, L., Li, X., Li, C., Zhang, H., Zheng, Y., Wang, J., Zhao, Y., Cheng, Q., Liu, C., Liu, S., Wang, X., 2014. Towards a common validation sample set for global land-cover mapping. Int. J. Remote Sens. 35, 4795–4814.
Zhong, L., Gong, P., Biging, G.S., 2014. Efficient corn and soybean mapping with temporal extendability: a multi-year experiment using Landsat imagery. Remote Sens. Environ. 140, 1.
Zhou, Y., Xiao, X., Qin, Y., Dong, J., Zhang, G., Kou, W., Jin, C., Wang, J., Li, X., 2016. Mapping paddy rice planting area in rice-wetland coexistent areas through analysis of Landsat 8 OLI and MODIS images. Int. J. Appl. Earth Obser. Geoinform. 46, 1–12.
Zucca, C., Wu, W., Dessena, L., Mulas, M., 2015. Assessing the effectiveness of land restoration interventions in dry lands by multitemporal remote sensing—a case study in ouled DLIM (Marrakech, Morocco). Land Degrad. Dev. 26, 80–91.