

Automated Cropland Mapping of Continental Africa using Google Earth Engine Cloud Computing

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

^aU. S. Geological Survey (USGS), 2255, N. Gemini Drive, Flagstaff, AZ 86001, USA

^bInternational Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, Hyderabad, India

^cBay Area Environmental Research Institute (BAERI), 596 1st St West Sonoma, CA 95476, USA

^dUniversity of New Hampshire, NH, USA

^eGoogle, MountainView, CA, USA

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 (sub-meter 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-meter 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 croplands areas (260 Mha cultivated plus 36 Mha fallows) and 330 Mha of gross cropland areas. Of the 260 Mha of croplands 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 rapidly, but currently, insufficient reference data exist to rigorously report trends from these results.

Key words: Cropland mapping, Classification, MODIS, Remote sensing products, Google earth engine, Africa,

List of Abbreviations

AVHRR Advanced Very High resolution radiometer

FAO Food and Agricultural Organization of the United Nations

*Corresponding Author

Email address: jxiong@usgs.gov (Jun Xiong)

Preprint submitted to Elsevier

- 15 FROMGC 30 m global cropland extent derived through multisource data
- 16 GCEV1 global cropland extent version 1
- 17 GFSAD250 Global Food Security Support Analysis Data (GFSAD) Cropland Products of Africa at 250-m resolution
- 18 GLC2000 global land cover for the nominal year 2000
- 19 GRIPC global rainfed, irrigated, and paddy croplands
- 20 LULC2000 land use land cover for the nominal year 2000
- 21 MCD12Q1 MODIS Land Cover Type product
- 22 MERIS MEdium Resolution Imaging Spectrometer
- 23 MODIS Moderate Resolution Imaging Spectroradiometer
- 24 SPOT Satellite Pour l'Observation de la Terre

25 1. Introduction

26 The extent, distribution, and characteristics (e.g., irrigation *versus* rainfed, cropping intensity) of croplands are factors that
 27 have long been identified as fundamental influences on agricultural development pathways, food security scenarios, and
 28 poverty reduction (Jayne et al., 2014). Estimates show that 52% of the world's remaining arable land is in Africa, yet
 29 most of this land is concentrated in just eight countries (Algeria, Democratic Republic of the Congo, Ethiopia, Morocco,
 30 Nigeria, South Africa, Sudan, Uganda), while a number of the remaining countries contain large rural populations clustered
 31 in remarkably small areas (Chamberlin et al., 2014). Demography of Africa is projected to change exponentially, where the
 32 population is expected to increase from the current 1.2 billion to nearly 4 billion by the end of the century (Gerland et al.,
 33 2014). A quarter of the population is undernourished and many countries experience famines in sub-Saharan Africa (Clover,
 34 2010). In this context, timely and dependable information on agricultural croplands of Africa is a prerequisite necessity to (i)
 35 isolate the agricultural croplands to assess crop water use, crop productivity, and crop water productivity, and (ii) investigate
 36 how the croplands respond to different climatic conditions(Waldner et al., 2015).

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

52 Further, there are several cropland studies. Time-series remotely sensed data are established as effective tool in cropland
 53 mapping (Esch et al., 2014) and have been successfully implemented at regional-scale (Bégué et al., 2014; Ding et al., 2014;
 54 Gumma et al., 2014; Helmholz et al., 2014; Pardhasaradhi Teluguntla et al., 2015) as well as at global scale (Chen et al.,
 55 2015; Pittman et al., 2010; Radoux et al., 2014; Salmon et al., 2015; Thenkabail and Wu, 2012; Wang et al., 2015). Various
 56 aspects of croplands are mapped such as irrigated areas (Conrad et al., 2016; Peña-Arancibia et al., 2016; Salmon et al.,
 57 2015; Thenkabail and Wu, 2012), rainfed areas (Biradar et al., 2009; Salmon et al., 2015), cropping intensities (Qiu et al.,
 58 2014), and crop types (Gumma et al., 2014; Zhang et al., 2015; Zhong et al., 2014; Zhou et al., 2016), and cropland fallows
 59 (Müller et al., 2015). There are many methods and techniques adopted for cropland classification that include phenology
 60 based algorithms (Dong et al., 2015; Jeganathan et al., 2014; Pan et al., 2015), classification regression trees (Deng and

61 Wu, 2013; Egorov et al., 2015; Ozdogan and Gutman, 2008), decision tree algorithms (Friedl and Brodley, 1997; Shao and
62 Lunetta, 2012), Fourier harmonic analysis (Zhang et al., 2015), spectral matching techniques (Dheeravath et al., 2010),
63 support vector machines (Mountrakis et al., 2011), random forest algorithm (Tatsumi et al., 2015) and a number of other
64 machine learning algorithms (DeFries, 2000; Duro et al., 2012; Lary et al., 2016; Pantazi et al., 2016). Many studies adopted
65 supervised and unsupervised classification approaches. Supervised methods (Egorov et al., 2015) rely extensively on *in-situ*
66 data or on human interpretation of spectral signatures, making the classification process resource-intensive, time-consuming,
67 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
68 case in Africa, supervised approaches lead to uncertainties. Unsupervised approaches require far less *in-situ* data or human
69 interpretation but they require large volumes of *in-situ* data for class identification and validation data.

70 Specific to continental Africa, amongst existing cropland products there has been large disagreement (Fritz and See, 2008;
71 Giri et al., 2005; Hansen and Reed, 2010; Herold et al., 2008; McCallum et al., 2006) especially in the extent of the cultivated
72 areas and their spatial distribution (Fritz, See, et al., 2011; Salmon et al., 2015; P Teluguntla et al., 2015; Thenkabail and
73 Wu, 2012; Waldner et al., 2015) as a result of fragmented and heterogeneous rural landscapes (Lobell and Asner, 2004) and
74 low agricultural intensification (Pittman et al., 2010) throughout continental Africa. The challenges of mapping cropland in
75 Africa also include: (a) spatial structure of the agricultural landscape (Vancutsem et al., 2012), (b) spectral similarity with
76 grassland, mainly in arid and semi-arid areas (Herold et al., 2006, McCallum et al. (2006)), (c) high regional variability in
77 terms of agricultural systems and calendars between the hyper-arid Sahara and other agro-ecological zones (Vintrou et al.,
78 2012).

79 Further, the current state-of-art using the above methods and approaches is mostly limited to producing cropland products
80 for a given period, or for a growing season, or for a particular year. However, such a process over very large areas such
81 as continent will always have limitations in availability of extensive collection of reference data. The biggest difficulty in
82 cropland mapping is in the lack of algorithms that accurately reproduce cropland products year after year or season after
83 season. So, more recently, there are efforts at producing cropland products by developing automated algorithms (Jamali et al.,
84 2014; Waldner et al., 2015; Yan and Roy, 2014). Thenkabail et al developed rule-based ensemble decision-tree Automated
85 algorithms to produce cropland *versus* non-croplands across years for Australia, Tajikistan and California (Teluguntla et
86 al., 2016; Thenkabail and Wu, 2012; Wu et al., 2014). Waldner et al. (2015) used a baseline map generated from five
87 knowledge-based temporal features to train an automated support vector machines (SVM) classifier on selected areas in
88 Argentina, Belgium, Ukraine, and China. However, these automated algorithms are currently applied only to small pilot
89 studies and not over large areal extent such as the African Continent.

90 Given the above discussions, the overarching goal of this research was to develop and test automated cropland mapping
91 algorithms (ACMAs) over a very large area such as a continent with an ability to automatically and accurately reproduce
92 cropland products year after year and season after season using MODIS 250-m 16-day time-series data. Africa was
93 chosen given its importance for food security in the twenty-first century and to address the challenge of mapping complex
94 agricultural systems. The spatial, temporal, and spectral specifications of MODIS are considered as highly suitable for
95 land use and land cover (LULC) classifications, especially for cropland extent and area mapping (Hentze et al., 2016). The
96 Google Earth Engine (GEE) cloud computing platform was used in this project to generate the products as well as collecting
97 reference data. The GEE is a system designed to enable petabyte-scale, scientific analysis and visualization of geospatial
98 datasets. Earth Engine provides a consolidated environment including a massive data catalog co-located with thousands
99 of computers for analysis. The user-friendly front-end provides a workbench environment to allow interactive data and
100 algorithm development, and support for in-the-field activities such as validation, ground-sampling, and crowd-sourcing.
101 We first, develop ensemble decision-tree algorithm ACMA for the year 2014 for the African continent, and then tested
102 and validated ACMA for the same year. This was followed by validation of ACMA algorithm for 11 independent years
103 (2003-2013). Finally, we deploy the ACMA algorithm on Google Earth Engine (GEE) cloud computing platform, so
104 scientists and practitioners can routinely reproduce cropland products of Africa year after year.

105 2. Data

106 2.1. Study Area

107 The study area included the entire African continent which extends from approximately 38°N to 35°S latitude, occupies
108 30.3 million km^2 , and has several distinct geologic and biogeographic regions with varying land cover types. For example,
109 Sahara, the largest hot desert in the world, comprises much of the land found within North Africa, excluding the fertile

110 coastal region situated against the Mediterranean Sea, the Atlas Mountains of the Maghreb, and the Nile Valley of Egypt and
 111 Sudan. Savannas, or grasslands, cover almost half of Africa, more than 13 million km^2 . These grasslands make up most
 112 of central Africa, beginning south of the Sahara and the Sahel and ending north of the continent's southern tip. Also, 80
 113 percent of Africa's rain forest is concentrated in central Africa, along the Congo River basin. Swahili Coast, stretches about
 114 1,610 kilometers along the Indian Ocean, from Somalia to Mozambique, where vegetated areas are located on a narrow
 115 strip just inland from the coastal sands and heavy cultivation has diminished the diversity of plant species in this interior
 116 area. Southern Africa will be one of the regions in the world whose crop production is most affected by climate change
 117 such as higher temperatures and reduced water supplies, along with other factors like biodiversity loss and ecosystems
 118 degradation (Lobell et al., 2008). All the raster and vector data in entire Africa continent were produced in Geographic
 119 projection (WGS84) at a spatial resolution of 0.0022458 degrees (equivalent to 250 m at the equator).

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

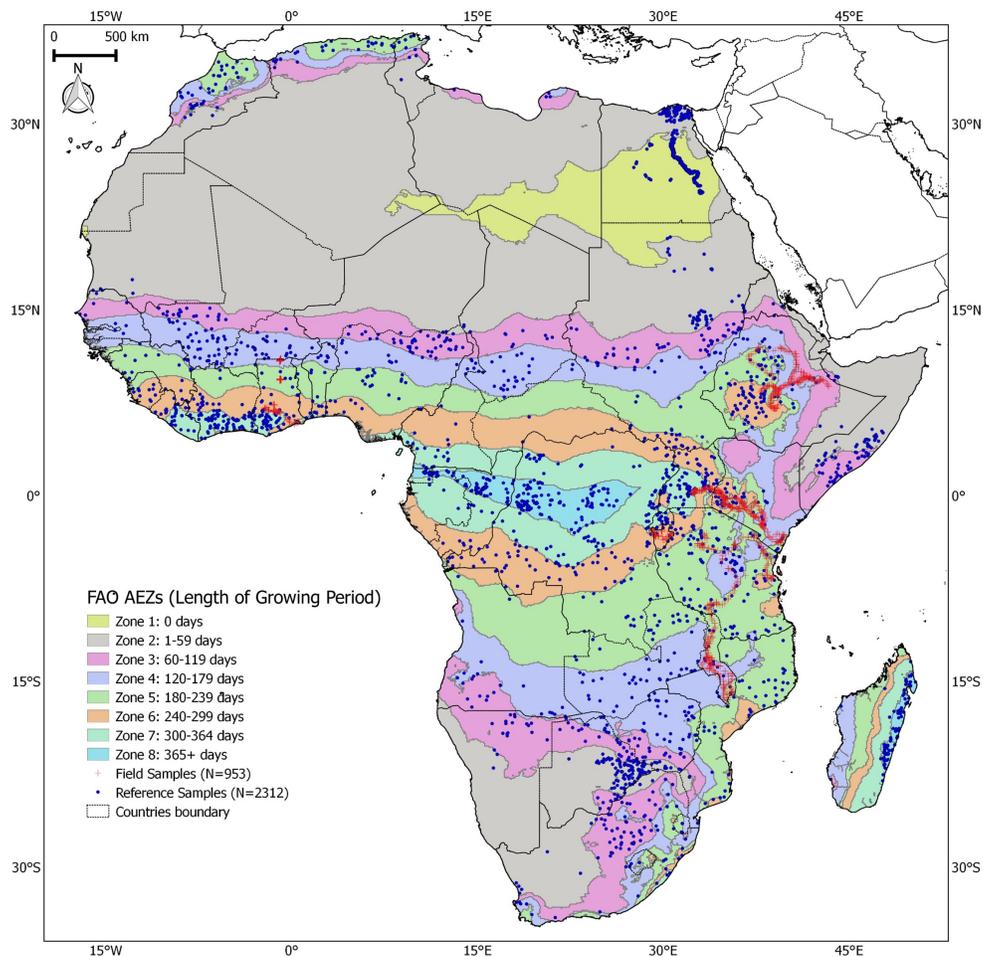


Figure 1: The United Nations (UN) Food and Agriculture Organization (FAO) Global Agro-Ecological Zones (AEZs) and distribution of reference samples repository in Africa continent. [Note: initial 15 AEZs were consolidated to final 8 to eliminate AEZs with zero or insignificant agriculture such as in the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

123 **2.2. Existing cropland/LULC reference maps**

124 Available land use/land cover (LULC) reference maps of Africa (Table 1) from different sources vary widely in how they are
 125 defined, derived, and mapped using a wide range of data, and methods and have different projections, formats, resolutions,
 126 and LULC categories. Even though they are used widely in LULC research, inconsistencies and uncertainties make their use

Table 1: Datasets used in creating 250m 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	300m	2005, 2009	LULC	(Arino, 2010)
Africover	FAO	Landsat 7	30m	1995-2002	LULC	(Kalensky, 2014)
LULC 2000	USGS	AVHRR	2000m	2000	LULC	(Soulard, 2014)
GLC 2000	JRC	SPOT	1/112°	2000	LULC	(Fritz et al., 2000)
MCD12Q1	NASA	MODIS	500m	2004 - now	LULC	(Leroux et al., 2014)
MODIS-JRC	JRC/MARS	MODIS, Landsat	250m	2009	LULC	(Vancutsem et al., 2013)
GCEV1	USGS	MODIS, Landsat	1000m	2010	Cropland	(Teluguntla et al., 2015)
Global30	NGCC	Landsat 7	30m	2010	LULC	(Chen et al., 2015)
FROMGC	CESS	Landsat 7	30m	Circa 2010	LULC	(Gong et al., 2013)
GRIPC	BU	MODIS	500m	Circa 2005	Cropland	(Salmon et al., 2015)

127 as reliable baseline maps questionable (P Teluguntla et al., 2015). Thereby, we expanded our investigation of the cropland
 128 extent by incorporating these studies into a comprehensive baseline crop layer (Figure 3) in Section 3.2.

129 2.3. Reference samples repository

130 In-situ samples collected from ground data is always the first step to establishing knowledge for the classifier in classification.
 131 These *in-situ* data are supposed to provide the most accurate information by definition. However, they are often not an
 132 ideal gold standard but degraded by error (Foody, 2010) because of small samples size, sampling bias and inconsistent
 133 labeling. A web-based system for supporting classification have been used in the past for general land cover (Fritz et al.,
 134 2009; Tsendbazar et al., 2015). In this paper, web-based data was developed (in addition to extensive ground data) using
 135 multiple sources and consolidated on the GEE platform.

136 The reference samples repository consists of following components: The project developed a ground data-collection mobile
 137 app <https://croplands.org/mobile> that can be downloaded and run on a smart phone device. This mobile app allows users
 138 to collect geo-references ground data that includes digital photos, cropland data required for the project (cropland *versus*
 139 cropland fallows, irrigated *versus* rainfed, cropping intensity). All data so collected from anywhere by anyone in the
 140 world is automatically uploaded to the project server. All data so collected from anywhere in the world is automatically
 141 uploaded to the project server. All data samples, so collected are further reviewed in the online image-interpretation tool
 142 (<https://croplands.org/app/data/classify>) to ensure that the samples are centered on the farm field using sub-meter to 5-meter
 143 very high-resolution imagery (VHRI) data from sensors such as Worldview 2, QuickBird, and IKONOS. Reference ground
 144 data for Africa were collected through several field campaigns by the project team in May, June, and August 2014 to
 145 coincide with the peak cropping seasons in different parts of Africa. Field information was collected from 250 m × 250
 146 m homogeneous plots. A total of 1381 samples were collected from Ethiopia, Ghana, Kenya, Malawi, Mali and Uganda.
 147 Reference data was also collected from several other sources. First, some other global/region projects (Tateishi et al.,
 148 2014; Zhao et al., 2014) shared with us valuable reference datasets. To incorporate them into our project, we converted
 149 their labeling system to be consistent with the labeling scheme of our project. Second, ~500 reference cropland samples
 150 were selected from a series of published literature for selected areas of Africa based on detailed studies using VHRI or
 151 high-resolution imagery such as Landsat (Haack et al., 2014; Kidane et al., 2012; Rembold et al., 2000; Shalaby and Tateishi,
 152 2007; Were et al., 2013; Zucca et al., 2015).

153 Overall, there were total 3,265 reference samples (Figure 1) spread across the eight consolidated agro-ecological zones
 154 (AEZs) of the African continent. Of these 953 reference samples were collected during the field visit by the team and the rest
 155 2312 reference samples were sourced from partners/collaborators (Figure 1). When the full reference samples repository was
 156 established through above approaches, every sample was then marked into “training” and “validation” groups. A random
 157 70%-30% splitting of the 3265 were used to separate 2285 samples for “training” and the rest 953 for “validation”. The 3265

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

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

165 2.4. MODIS NDVI times series data

166 The MODIS 250m 16-day composite NDVI product was found to have high temporal resolutions to overcome the data
167 gap because of cloud cover and harmattan haze during the monsoon season over Africa (Leroux et al., 2014; Vintrou et
168 al., 2012). Hosted on Google Earth Engine (GEE), the MYD13 product is computed from daily atmospherically corrected
169 bi-directional surface reflectance that has been masked for water, clouds, heavy aerosols, and cloud shadows. Google Earth
170 Engine, based on millions of servers around the world and the state-of-the-art cloud-computing and storage capability, has
171 archived a large catalog of earth observation data and enabled the scientific community to work on the trillions of images in
172 parallel processing way (Hansen et al., 2013).

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

179 3. Methodology

180 3.1. Method overview

181 The proposed methodology is presented in Figure 2. First, MODIS 16-day 250 NDVI imagery composite of the African
182 continent was stratified by (1) Masking out the non-cropland area using 250m baseline cropland mask of Africa (Figure 3),
183 (2) Sub-setting masked area into eight consolidated FAO agro-ecological zones (AEZs) (Figure 1), and (3) clustering each
184 cropland mask of the 8 AEZs into 25 unique clusters using K-means algorithm for a total of 200 classes. Second, ground
185 samples from reference samples' repository (section 2.3) were split into training part and validation part. The former was
186 used to characterize unique ideal time-series signatures. Third, clustered classes from each of the 8 AEZs generated using
187 k-means algorithm were grouped together through quantitative spectral matching techniques (QSMTs) and the group of
188 similar cluster classes was matched with the ideal spectra to identify and label classes. The class labeling is further verified
189 through ground data, VHRI, field visits, and through external sources. The process lead to an accurate reference cropland
190 layer (RCL) of Africa for the year 2014 (RCL2014). The cropland knowledge available in the RCL2014 was then coded
191 in an ensemble decision-tree automated cropland mapping algorithm (ACMA) to accurately replicate cropland products
192 through Automated cropland layer for the year 2014 (ACL2014). Once this is achieved accurately, ACMA was deployed on
193 the Google Earth Engine to create automated cropland layers for independent years from 2003 to 2013.

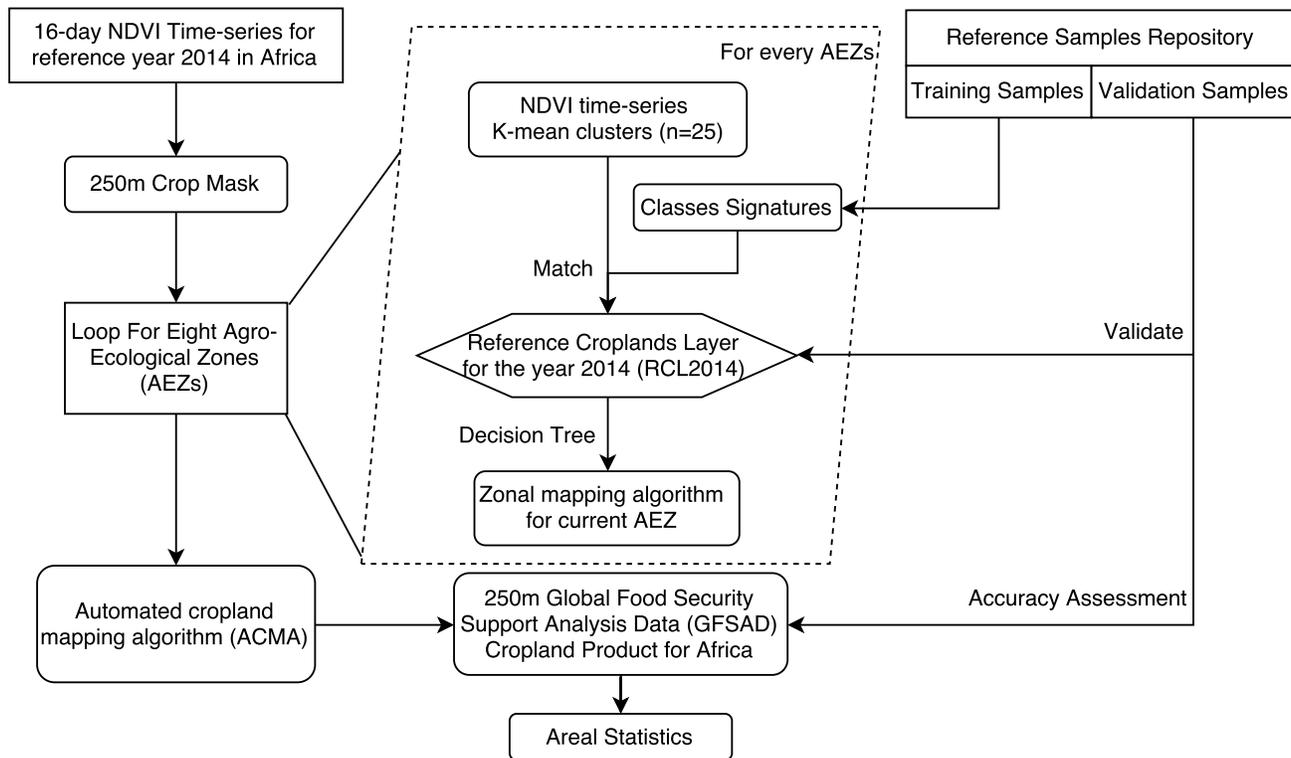


Figure 2: Schematic diagram of the methodology and area statistics used in creating reference cropland layer for the year 2014 (RCL2014) and the Automated cropland mapping algorithm (ACMA) for continental Africa.

194 3.2. 250-m Baseline crop mask

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

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

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

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

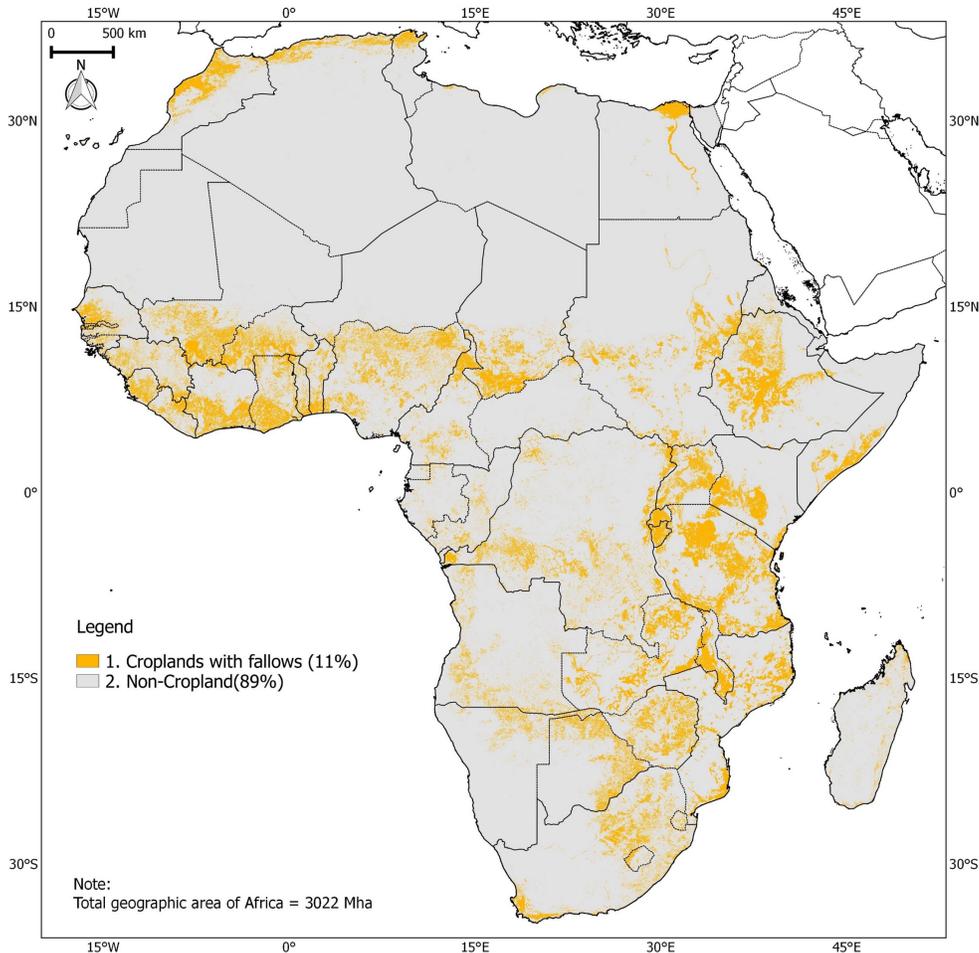


Figure 3: Baseline cropland mask of Africa based on 10 pathfinding studies. Aggregated 250m cropland mask derived from 10 previous studies in Africa, including cropland fallow areas (Table 1).

217 3.3. Classification System and signatures

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

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

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 250m pixel (~6.25 ha).

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

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

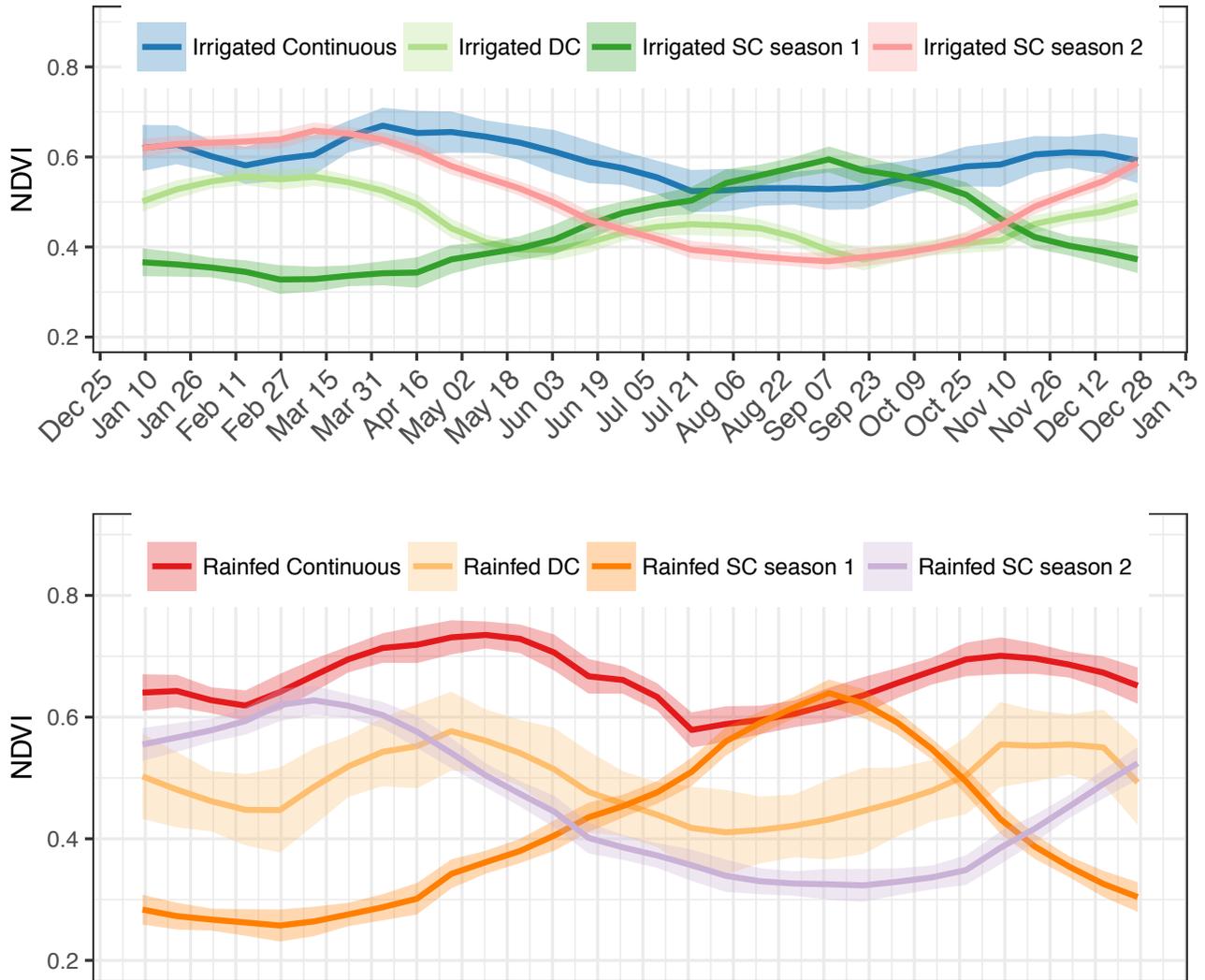


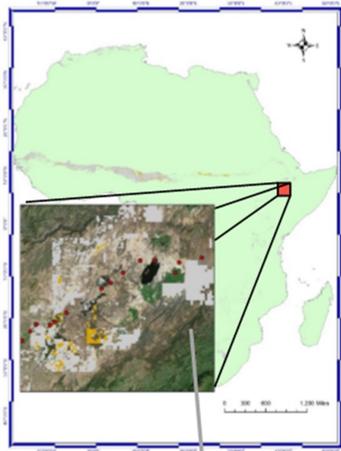
Figure 4: Ideal spectral signatures of the distinctly separable, unique four irrigated (top) and four rainfed (bottom) classes in agro-ecological zone 3 (AEZ 3), Africa. Illustration on the every 16-day time-series of MODIS 250 m NDVI profiles based on ground data sample knowledge base collected throughout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

249 **3.4. Creation of Reference Cropland Layer (RCL)**

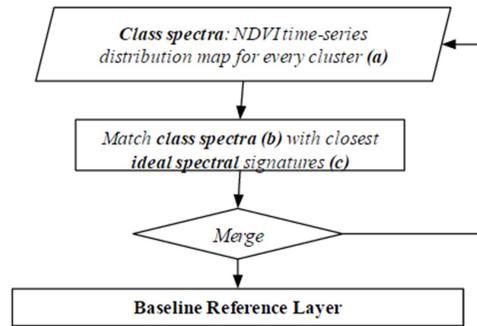
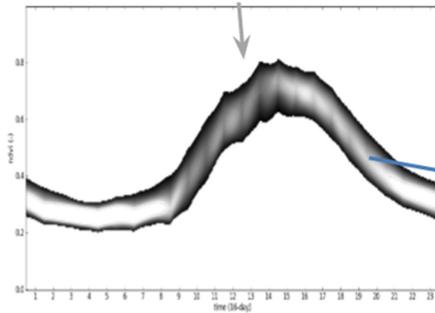
250 To drive clustering to the massive dataset on terabyte level, all the MODIS tiles covering Africa in 2014 were organized as a
 251 large ImageCollection in Google Earth Engine and then exported in parallel netCDF format (PnetCDF, Li et al. (2003)) on
 252 NASA Earth Exchange (NEX, Nemani et al. (2011)) supercomputing platform. Message Passing Interface (MPI) k-means
 253 (Zhang et al., 2011) algorithm was applied to do the clustering with 2,000 CPUs on NASA AMES super computer. For
 254 the total 8 AEZs, K-means cluster results in a total 200 unique clusters for the continental Africa, based on their NDVI
 255 time-series profile signal.

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

(a) Cluster result for Class 5 Zone 3



(b) Class spectra: Distribution of the MODIS NDVI distribution map for the generic class 5, in Zone 3. <1% of the noisy points removed



(c) Ideal Crop Signature for rainfed croplands in Zone 3 : Matching with of "rainfed, single, season 1" signatures

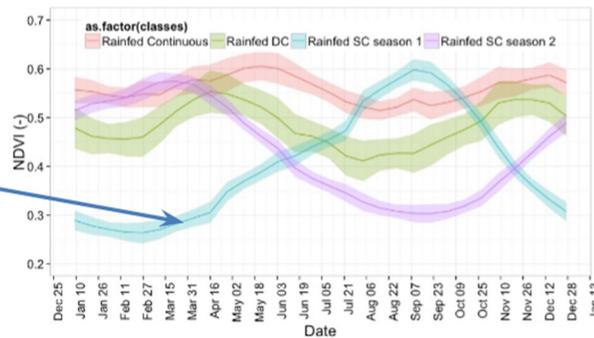


Figure 5: Quantitative Spectral matching (QSM) of a generic class with an ideal spectral signatures.

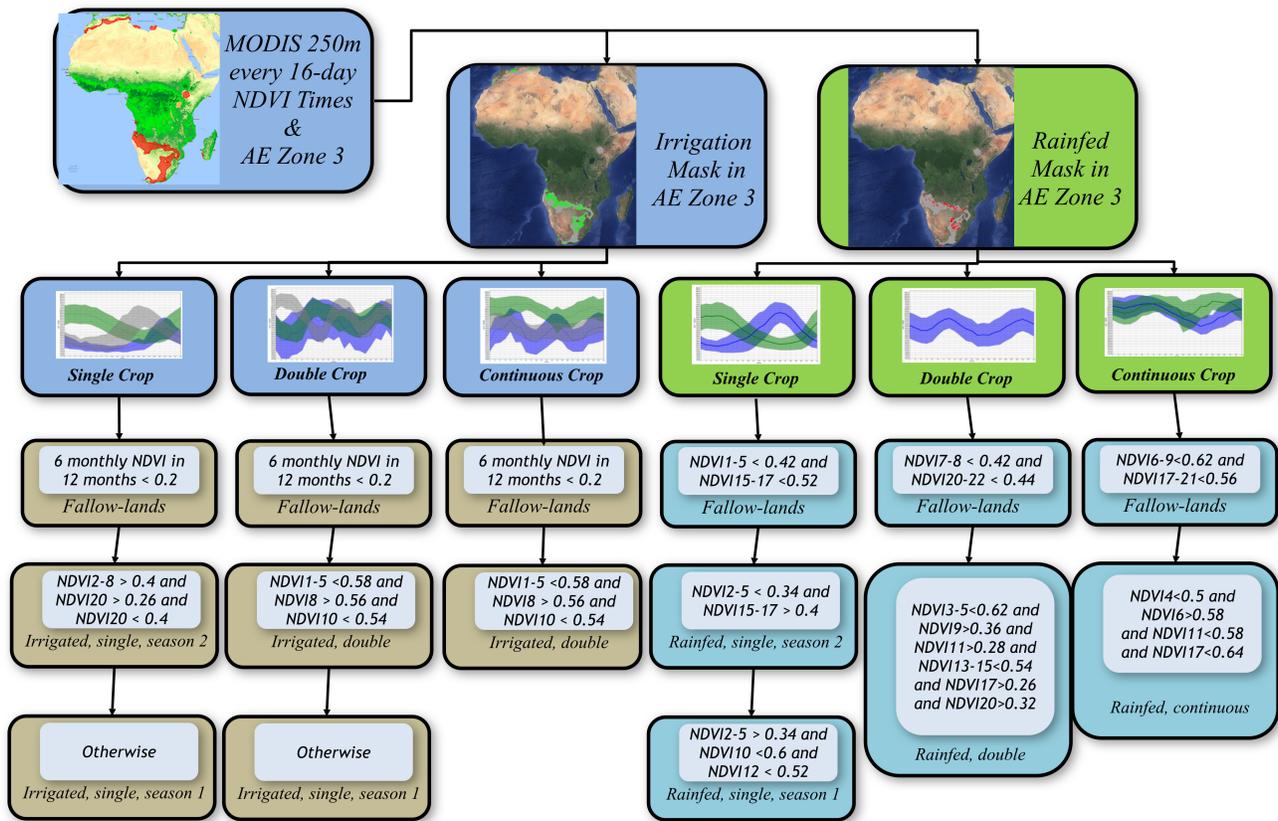
261 3.5. Generalization of RCL to ACMA Rules

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

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

276 An example to distinguish MODIS NDVI time series distribution of these classes in Africa are shown in Figure 4 for AEZ3.
 277 Similar eight classes were established across all AEZs. Apart from the eight distinct classes (Figure 4) across AEZs, a
 278 cropland fallow class is also coded based on NDVI falling below a threshold during the critical growing period. The nine
 279 classes (Table 2) from the irrigated and rainfed masks of the 8 AEZs are analyzed (e.g., Figure 5) leading to RCL2014. The

280 knowledge captured in the 9 RCL2014 classes are then coded in ACMA (Figure 6) to derive ACMA developed cropland
 281 layer for the year 2014 (ACL2014). The process of developing the ACMA go through numerous iterations, as illustrated
 282 partially in Figure 6. It involves writing a bunch of simple rules to capture RCL2014 knowledge in the codes and replicate it
 283 accurately. Every ACMA rule captures certain percentage of total cropland area and its characteristics (e.g., irrigated *versus*
 284 rainfed or intensity) in each of the nine classes (Table 2) of RCL2014. The process is repeated with numerous additional rules
 285 to capture as much cropland area/extent and as many cropland characteristics as possible. If the rule captures non-croplands
 286 then the iteration is repeated by tweaking the rule till we can precisely (or near precisely) capture croplands, distinguish them
 287 from non-croplands, as well as differentiate irrigated croplands from rainfed croplands or cropping intensities. The process
 288 requires several runs to slightly adjust and re-adjust the thresholds till ACL2014 achieves as close a match as possible with
 289 RCL2014. The ACMA rules are shown in Figure 6.



Note: NDVI 1-5 represents bands from 1st to 5th of the total 23 16-day NDVI bands in one calendar year

Figure 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.

290 3.6. Ensemble and Deployment algorithm on Google Earth Engine for year-to-year-classification

291 ACMA is a group of decision-trees like what we show in Figure 6 so we can easily deploy it on Google Earth Engine and
 292 run fast for the independent years. Taking MODIS 250-m time-series data as input, we tested ACMA algorithm from 2003 and
 293 2013. This entire ACMA algorithm is made available here: http://geography.wr.usgs.gov/science/croplands/algorithms/africa_250m.html

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

296 through 2013 and tested the ACMA algorithm.

297 3.7. Areal Statistics

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

$$SPAs = FPAs \times CAF$$

301 Where CAFs are determined by taking an average MODIS NDVI image during the growing season and plotting all pixels
302 of the class for this period from the MODIS NDVI image in a brightness-greenness-wetness space (Thenkabail et al.,
303 2007). The same methodology is adopted here. Also, to get actual areas, one need to re-project MODIS cropland products
304 to appropriate projection. Further, areas are established during different seasons by accounting intensity (single, double,
305 triple, or continuous cropping). Areas cropped twice have areas counted two times a year. Single and continuous have areas
306 computed one time a year.

307 4. Results

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

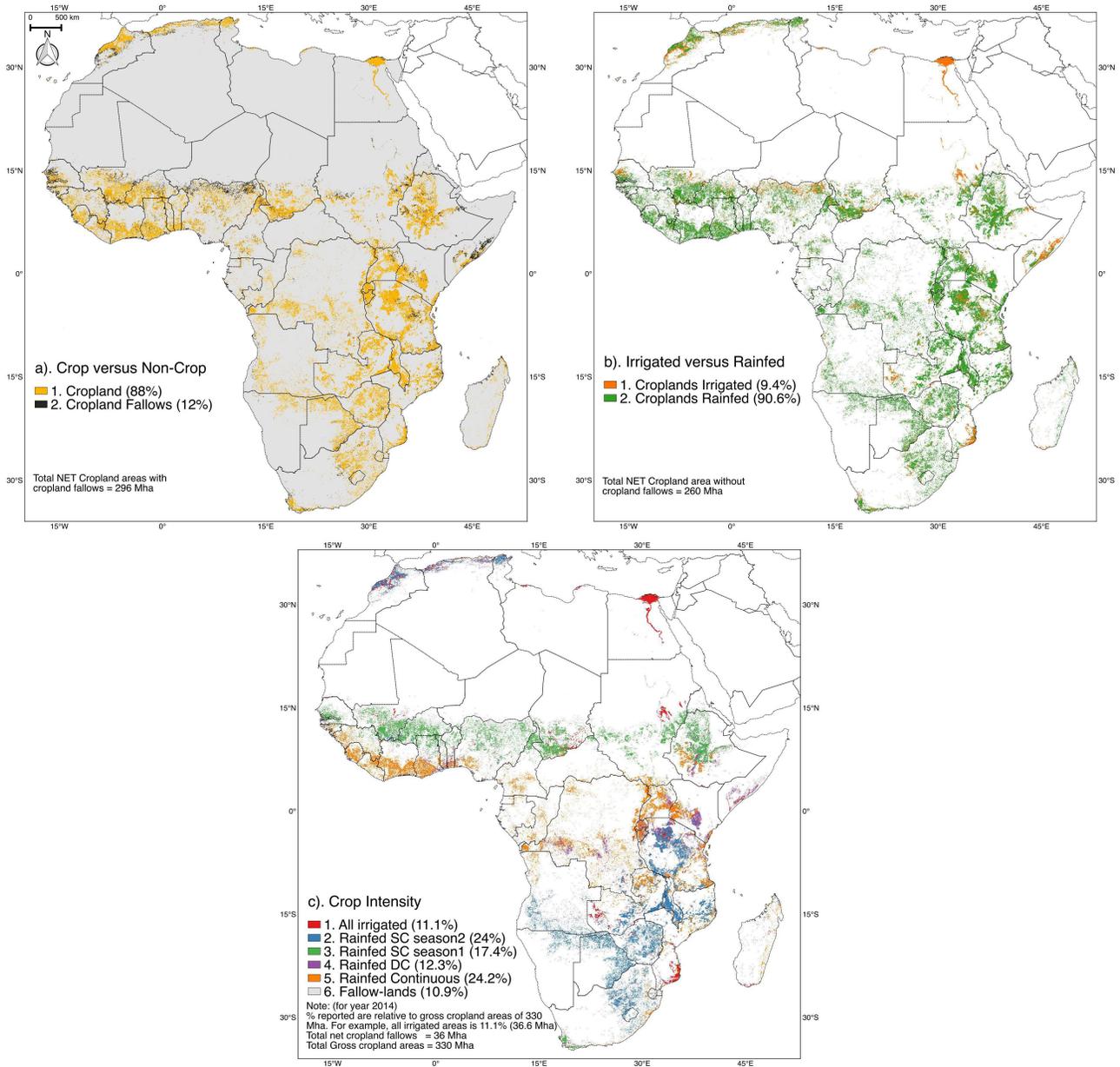


Figure 7: RCL2014 Spatial distribution. Reference cropland layers of Africa for the year 2014 (RCL2014). Three RCL2014 products:(a) Cropland *versus* non-Cropland Layer, (b) Irrigated *versus* Rainfed Layer, and (c) Crop intensities Layer.

312 4.1. Reference cropland layer of Africa for the year 2014 (RCL2014)

313 4.1.1. Croplands *versus* non-croplands

314 The accuracies of croplands *versus* non-croplands were evaluated for each of the 8 agro-ecological zones (AEZs) and the
 315 overall accuracies (OAs) varied between 89 to 100% (Table 3). The accuracy of the resulting cropland products was validated
 316 with the global food security support analysis data (GFSAD) project Validation Dataset <https://croplands.org/app/data/search>,
 317 which is a consistent global cropland validation dataset designed for validating cropland products and includes multiple
 318 datasets that are ground-based, VHRI based, or sourced from other local detailed studies. In this research a total of 3265
 319 samples, distributed over various agro-ecological zones (AEZs) of Africa, were collected through the crowdsourcing land
 320 cover validation tool called cropland.org. Also, the proposed product is compared with cropland statistics derived from other
 321 gridded and survey-based data sources. The AEZs 4, 5, 6, 7, and 3, together where about 95% of Africa’s croplands exist,

322 have an accuracy of 89-96%. The very high (98-100%) percent accuracies were for AEZs with very low (0.35 to 4.64%)
 323 cropland areas.

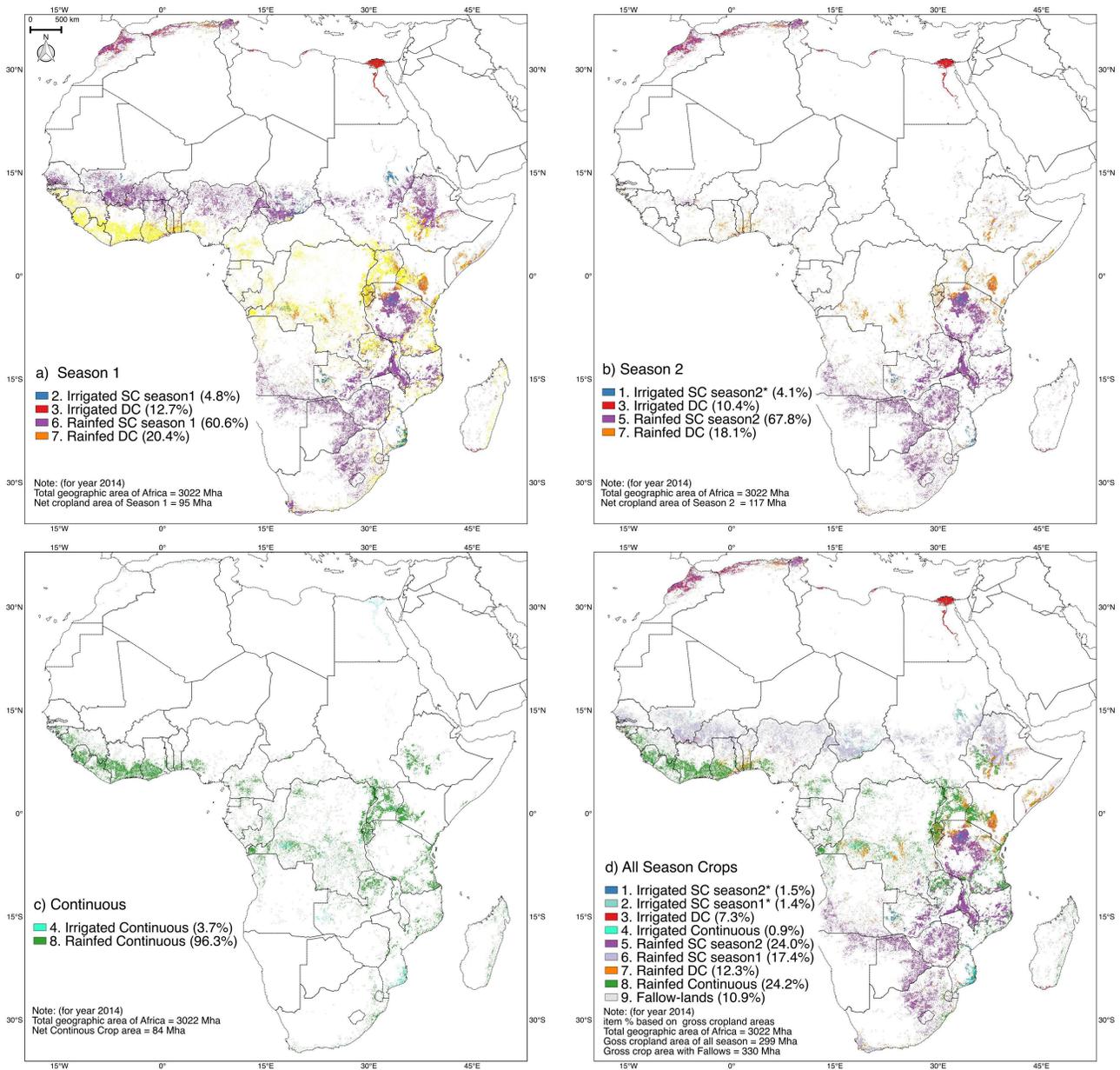


Figure 8: RCL2014 seasonal cropland layers. Reference cropland product of the year 2014 (RCL2014) for Africa at 250m generated using MODIS every 16-day time-series data, extensive field knowledge, image classification, and quantitative spectral matching techniques (QSMTs) methods. The top layer shows 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.

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

328 Besides, a country by country cropland areas was then computed and compared with MIRCA2000 ((Portmann et al., 2010),
 329 Figure 8; The most updated statistics were obtained through personal communication with Portmann and Siebert in 2014 to

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	% of total Cropland	Non-Cropland Area	% of total non-Crop Area	Crop Samples	Non-Crop Samples	Overall Accuracy
	Mha	%	Mha	%	-	-	%
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

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.

GCP250 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

330 coincide with our 2014 synthesis). The variability was maximum smaller island nations (e.g., Comoros, Mauritius). Few
 331 other countries (e.g., Sierra Leone, Cote dVoire, Chad, Guinea, and Cameroun) also showed significant variability. R-square
 332 of 0.42 is calculated based on all 55 African Nations (recognized by the United Nations and African Union) (Figure 9). If
 333 we 8 outlier countries, where uncertainty is maximum, the comparisons between: GFSAD250 with MIRCA2000 for the
 334 rest 47 countries increases to an R-square of 0.69. The GFSAD250 These results clearly imply the ability of GCEA250 to
 335 compute cropland areas of Africa and provide country level statistics.

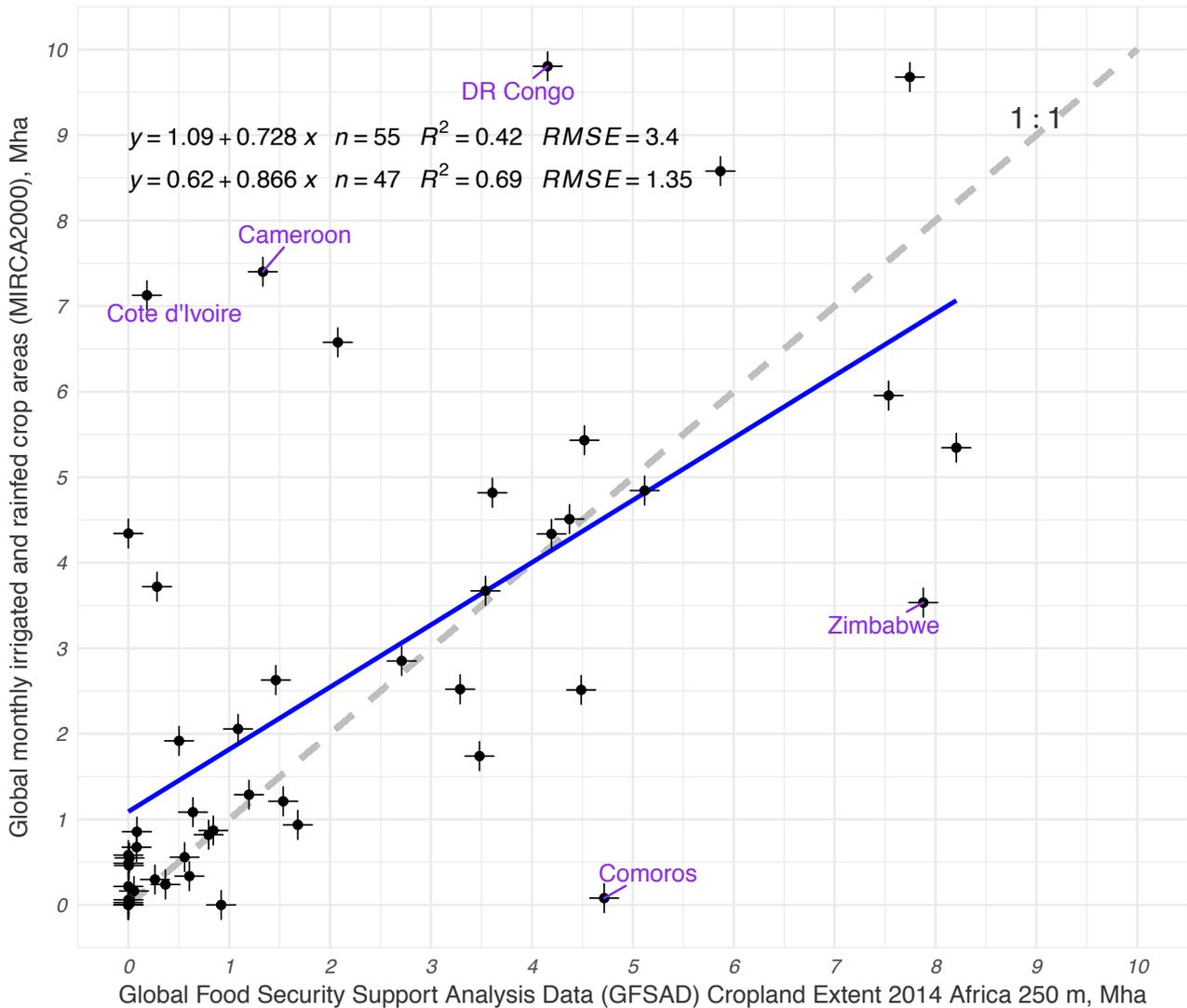


Figure 9: The global cropland product of Africa @ 250-m (this study) or GFSAD250 derived country by country cropland areas (rainfed+irrigated) of Africa compared with MIRCA2000 (Portmann, 2010).

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

- 339 1. Uncertainty in the calculation of MIRCA2000 areas. MIRCA2000 is a derived gridded dataset based on FAOSTAT
 340 database (Portmann et al., 2010). FAO compiles the statistics reported by individual countries, which are based
 341 on national censuses, agricultural samples, and questionnaire-based surveys with major agricultural producers,
 342 independent evaluations (FAO, 2006 and The World Bank, 2010). Since each country has its own mechanism,
 343 differences in data gathering, and resource limitation, the data lacks objectivity in many countries resulting in data

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

344 quality issues, particularly in Africa. For example, in 2008/09 in Malawi, cropland extent was estimated by combining
345 household surveys with field measurements derived from a “pacing method” in which the size of crop fields is
346 determined by the number of steps required to walk around them (Dorward and Chirwa, 2010).

347 2. Application of the ACMA over certain regions have to face the limitation of spatial resolution of MODIS pixels.
348 A typical case is Madagascar, in its slash-and-burn agriculture for pluvial rice which is a predominant component
349 in of cultivation. These fields are easily mixed with neighboring vegetation because lack of cropland management
350 (Messerli and Messerli, 2009), resulting in fallow re-growth in rice fields.

351 4.1.2. Irrigated versus rainfed croplands

352 Of the 260 Mha croplands during 2014, 90.6% (236 Mha) was rainfed and just 9.4% (24.5 Mha) was irrigated (Figure
353 7b). Africa has 15% of the world population, but just 6% of global irrigated area of 400 Mha (Thenkabail et al., 2009,
354 2012) is in Africa. An overwhelming proportion of the irrigated areas were along the Nile, specifically in Egypt, North
355 Africa, South Africa, along Niger in Mali, and scattered irrigated areas in Southern Africa especially Lake Victoria and Lake
356 Malawi (Figure 7b). Irrigated *versus* rainfed classification accuracies were evaluated in each of the 8 AEZs and the overall
357 accuracies were between 89 to 94% (Table 5). The accuracy of the irrigated *versus* rainfed cropland products was validated
358 with the global food security support analysis data (GFSAD) project Validation Dataset as discussed before. A country by
359 country comparison of the irrigated areas and the rainfed areas computed by this study with MIRCA2000 reported statistics
360 are plotted in Figure 10. The R-square values were 0.6 for irrigated areas and 0.31 for rainfed areas. Irrigated areas can be
361 computed with great certainty and uncertainties were greatest for the small island nations and few other countries. If the 8
362 outlier nations are removed, for the rest 47 countries the R-square values with MIRCA2000 increased to 0.6 for the rainfed
363 areas. As expected, uncertainties were higher for rainfed croplands and this was mainly as a result of highly fragmented, low
364 biomass croplands that were either confused with grasslands in savannas or regrowth vegetation in the humid tropics. It also
365 has to be noted that MIRCA2000 data which relays on the national statistics also has great degree of uncertainty in rainfed
366 cropland estimates.

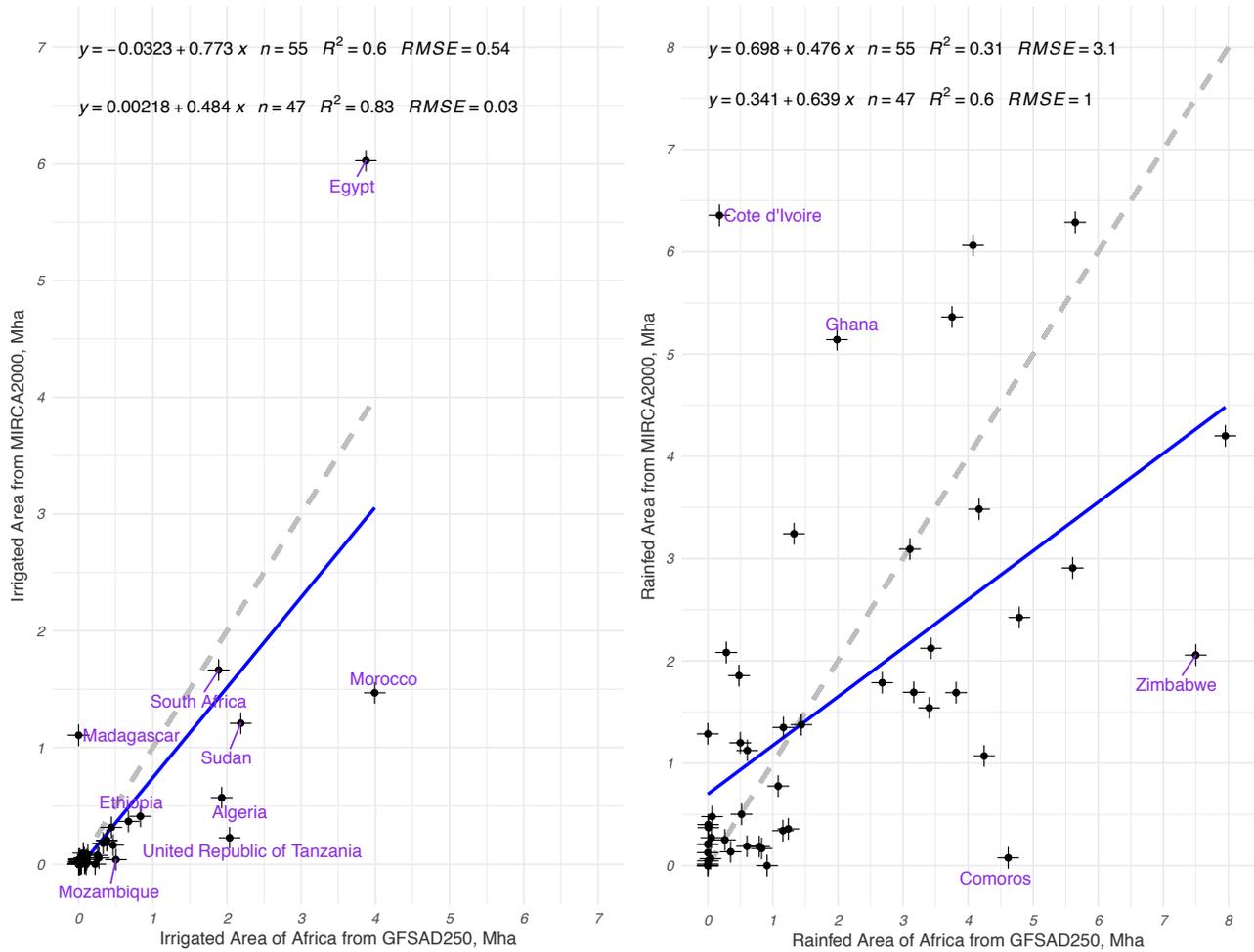


Figure 10: The global cropland product of Africa @ 250-m (this study) or GFSAD250 reference cropland layer for the year 2014 (RCL2014) irrigation/rainfed country area vs MIRCA2000. (a) Comparison of country-level estimates of cropland area from the new dataset presented in this paper against corresponding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

367 4.1.3. Cropping intensity

368 In both irrigated and rainfed areas of Africa, single crop is overwhelmingly dominant (Figure 7c, and Table 6). Of the 330
 369 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).
 370 This was followed by rainfed continuous crop gross areas with 79.83 Mha, rainfed double crop gross areas with 40.56 Mha,
 371 and irrigated double crop gross areas with 24.17 Mha (Table 6). Gross areas of irrigation single crop during season 1 (4.61
 372 Mha), season 2 (4.84 Mha), and continuous (3.07 Mha) were much smaller. Cropland fallows were 36 Mha during 2014,
 373 almost all of that in rainfed croplands with a negligible portion in irrigated croplands (Table 6).

374 4.1.4. Cropping seasonal Layer

375 Cropland areas are also mapped for two main seasons, continuous crops and a combination of the two seasons (Figure 8,
 376 Table 6). Season 1 (January-May) and season 2 (June-September). Much of the season 1 crops are in Southern Africa and
 377 North Africa, while season 2 is mainly distributed in West and Central Africa. Irrigated crops and continuous plantation
 378 crops are seen in both seasons, while continuous crops concentrated in West and Central Africa. Overall, for entire Africa,
 379 net cropland areas (NCAs) for season 1 was 95 Mha (Figure 8a, Table 6) and for season 2 was 117 Mha (Figure 8b, Table 6),
 380 and 84 Mha for continuous (Figure 8c).

Table 6: RCL2014 cropland area seasonal and total statistics of Africa for the year 2014 using MODIS 250m 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

381 4.2. Error matrix comparing ACP2014 with RCL2014

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

388 4.3. ACMA derived annual cropland layers from 2003-2014

389 We applied ACMA algorithm for 11 independent years (2003 through 2014) using MODIS 250 m every 16-day time-series
390 data of these years available on Google Earth Engine. The results as depicted in Figure 11 showed that the: a) net cropland
391 areas (NCAs) of Africa increased by about 11 Mha from 2003 to 2014, varying from 253 Mha to 264 Mha; b) gross cropland
392 areas (GCAs) Africa also increase by about 13 Mha from 2003 to 2014, varying from 323 Mha to 330 Mha; c) cropland
393 fallows of Africa decreased by about 10 Mha from 2003 to 2014, varying from 43 Mha to 30 Mha. This is, roughly an
394 increase of 1 Mha of croplands per year, whereas there was a decrease of 1 Mha of cropland fallows per year during the
395 same period. This can only increase further with rapid increase in population and increase food and nutritional demands of
396 the populations. The ability of ACMA algorithm to compute croplands as well as cropland fallows is important one. In
397 drought year cropland fallows increase and cropland areas decrease.

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

		RCL2014									
Class		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%
	2. Irrigated, SC, season 1	25,992	772,998	10,400							95.5%
	3. Irrigated, DC,		9,455	548,804	13,908						95.9%
	4. Irrigated, Continuous			11,122	514,613						97.9%
	5. Rainfed, SC, season 2					14,207,104	501,841				96.6%
	6. Rainfed, SC, season 1					731	12,564,794				96.1%
	7. Rainfed, DC,						145,763	4,570,981	581,352		86.3%
	8. Rainfed, Continuous							166,553	11,270,691	208,927	96.8%
	9. Fallow-lands								166,553	16,994,878	99.0%
	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									

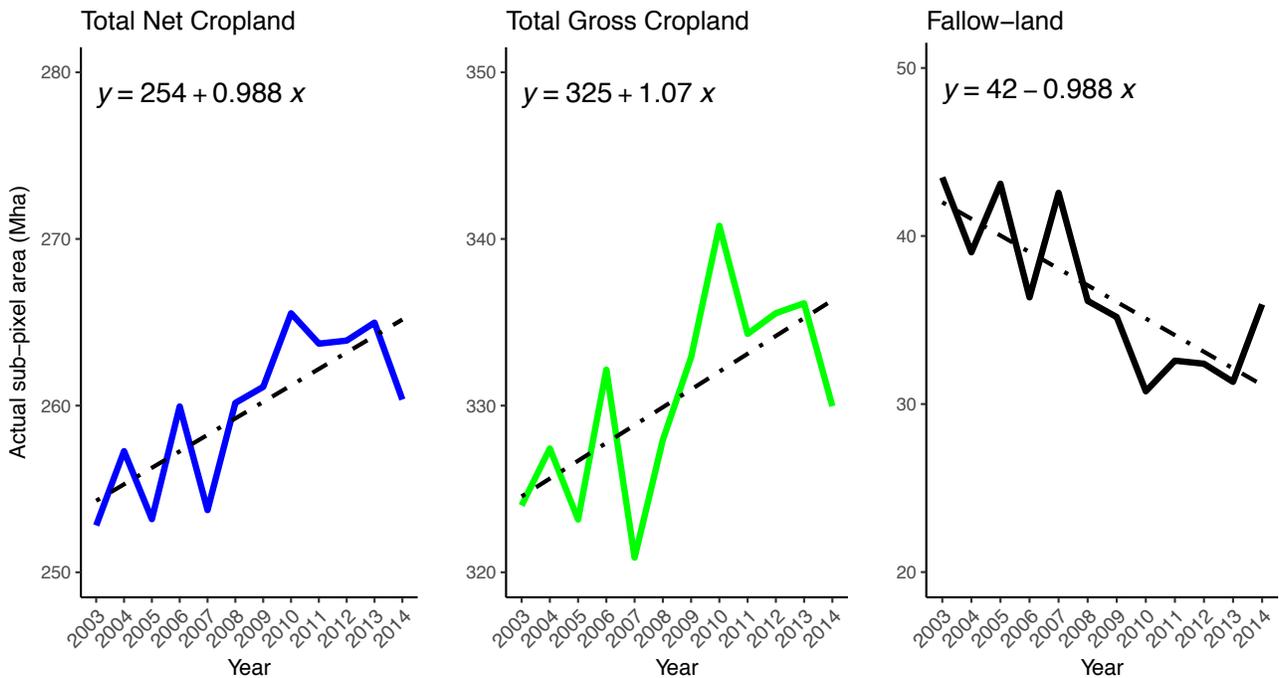


Figure 11: ACL2003 to ACL2014 derived cropland areas versus cropland fallow areas.

398 The ability of ACMA to capture variability are depicted between precipitation, ndvi and cropland areas. During drought
 399 years, we see dramatic: 1. increase in cropland fallows, and 2. decrease in cropland areas. Also during drought years, there
 400 is a significant decrease in the vigor of the existing croplands as illustrated by MODIS 250-m time-series NDVI plots. For
 401 example, in the 40,337 hectares' portion of area for AEZ3 depicted in Figure 13, 57% was cropland fallows during the
 402 drought year of 2005, whereas during the normal year of 2008 there was 35% fallow and during the good year of 2006 there
 403 were only 4% fallows. Similarly, in 0.62 Mha portion of cropland area in AEZ5 depicted in Figure 12, 21% was cropland
 404 fallows during the drought year of 2005, whereas during the normal year of 2008 there was 12% fallow and during the
 405 good year of 2006 there were only 10% fallows. The NDVI vigor trends also clearly depict drought, normal, and good
 406 years. Thereby, the ability of ACMA to highlight the combination of the above three factors highlights its value in assessing
 407 food security. There can not be direct accuracy assessment of other years without ground reference data. Nevertheless,
 408 We established an online utility called [CropRef](https://croplands.org) to generate reference samples using crowdsourcing in our project website
 409 <https://croplands.org>. This allows the use of Very High Resolution Imagery (VHRI) from DigitalGlobe and similar sub-meter

410 to 5-meter imagery to help generate year specific validation data in the future.

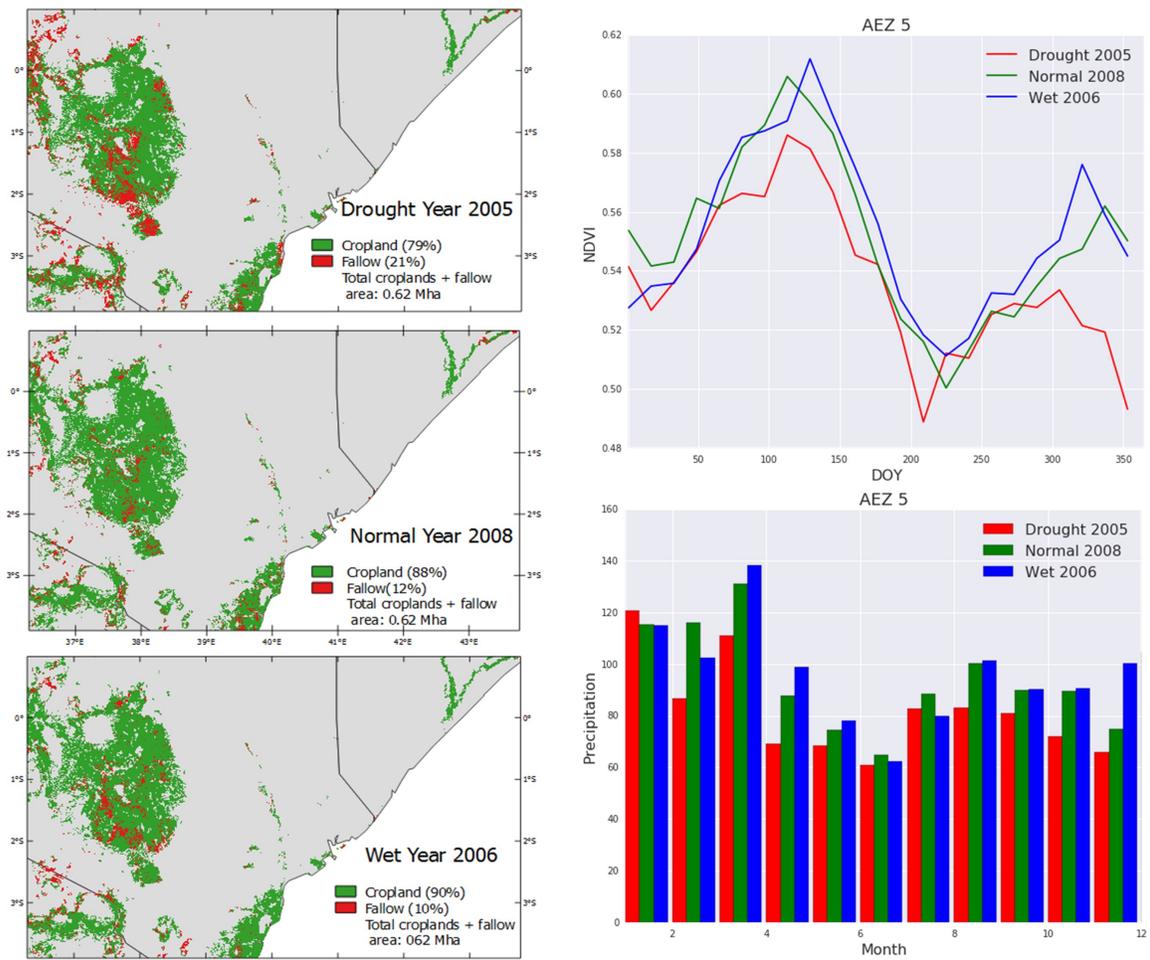


Figure 12: ACMA derived croplands versus cropland fallows for a drought year (2005), normal year (2008), and wet year (2006) in AEZ3. The figure shows spatial distribution of croplands versus cropland fallows (left), mean MODIS 250-m NDVI during the three-year (top right) and precipitation (bottom right).

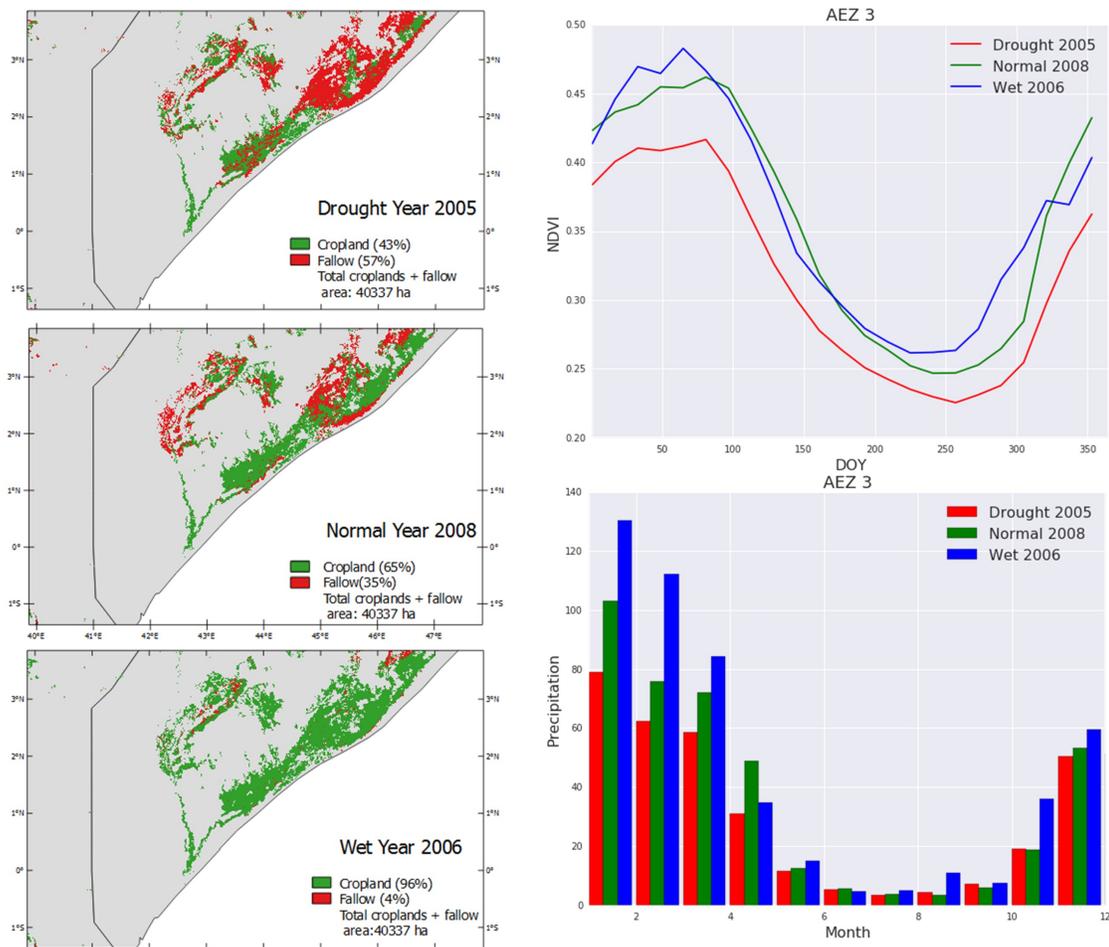


Figure 13: ACMA derived croplands *versus* cropland fallows for a drought year (2005), normal year (2008), and wet year (2006) in AEZ5. The figure shows spatial distribution of croplands *versus* cropland fallows (left), mean MODIS 250-m NDVI during the three-year (top right) and precipitation (bottom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

411 5. Discussion

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

425 To achieve greater accuracies, development of ACMA need to be done considering: (a) cropland masks, (b) AEZs, and (c)
 426 richness (quality, quantity, and spatial spread) of the reference data. The AEZs help us focus on certain agro-climatic zones
 427 and capture their unique characteristics. Along with the AEZ approach, the 250-m Crop mask derived from multiple sources

428 is an important starting point for this study to make significant advance from previous studies. This allowed us to conduct
429 this research by focusing heavily within the cropland mask, where majority (>95%) of the present croplands of Africa exists.
430 Nevertheless, it is important to check any expansion of croplands beyond the existing cropland mask. This requires us to 1)
431 carefully choose the way to stratify the input MODIS data; 2) Collect reference data using crowdsourcing technique and
432 interpret them with correctly. Until we can establish the effective knowledge-based decision-tree and verify the classification
433 output with acceptable accuracies, it is not prudent to apply recursive decision tree ACMA algorithm across the continent
434 with equal certainty. That is all the more reason to approach the ACMA development using AEZs.

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

444 The goal was not to map to many classes where achieving high accuracies becomes complicated, but replicating them year
445 after year (section 4.3) accurately becomes extremely difficult over very large areas. However, mapping a known number of
446 classes accurately and with ability to replicate year after year also accurately is crucial and meets the important challenge of
447 gathering routine and repetitive cropland statistics over time and space, thus contributing to food security (Figure 9, 10,
448 11). Often the knowledge of the zonal decision-trees that comes from the reference data sourced from ground samples,
449 photo-interpretation, and expert-knowledge for that zone enriches the recursive ACMA rules for that zone and extrapolation
450 of the same to other zones may not be applicable.

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

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

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

475 6. Conclusions

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

485 The ACMA generated cropland layer for the year 2014 for Africa (ACL2014) when validated showed overall accuracies
486 greater than 89% for each of the eight AEZs. This demonstrated the ability of ACMA to automatically produce cropland
487 products with acceptable accuracies. A country-by-country cropland areas statistics of all 55 African Countries generated
488 from this study was compared with the national census data based MIRCA2000 which were also updated in the year 2014.
489 The relationships showed significant correlations with R-square values between 0.6 to 0.83 for 47 of 55 countries. A
490 pixel-based agreement between the map produced in this study and a number of other studies showed uncertainties varying
491 between 15 to 25%. Overall, for the year 2014, the net cultivated cropland area for the entire African continent was 260
492 Mha with an additional 36 Mha left fallow. The gross cropland area was 330 Mha. Of the 296 Mha (cultivated + fallow) of
493 cropland areas, 91.7% (271 Mha) was rainfed and 8.3% (25 Mha) was irrigated. Net cropland area distribution in Africa
494 was 94 Mha during season 1, 117 Mha during season 2, and 84 Mha continuous.

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

506 Acknowledgements

507 The authors would like to thank following persons for their support: Dr. Felix T. Portman and Dr. Stefan Siebert for
508 providing statistics of MIRCA2000 ((Portmann et al., 2010), also latest statistics through personal communication between
509 Dr. Stefan Siebert and Prasad S. Thenkabail); Dr. Peng Gong for sharing of FROMGLC Validation Dataset (Zhao et al.,
510 2014); Dr. Ryutaro Tateishi for sharing of CEReS Gaia validation data Tateishi et al. (2014), and Dr. Friedl Mark for
511 sharing GRIPC500 dataset for inter-comparison. Special thanks to Dr. Fabio Grita and Dr. Michela Marinelli's help of
512 FAO/CountrySTAT team. Jennifer Dungan (NASA Ames Research Center) and Mutlu Ozdogan (University of Wisconsin-
513 Madison) area acknowledged for reviewing an earlier version of the manuscript. This study was supported by the NASA
514 MEaSUREs (Making Earth System Data Records for Use in Research Environments). The project is funded by NASA
515 MEaSUREs (NNH13AV82I) and the USGS Sales Order number is 29039. We gratefully acknowledge this support. The
516 United States Geological Survey (USGS) provided supplemental funding as well as numerous other direct and indirect
517 support through its Land Change Science (LCS), and Land Remote Sensing (LRS) programs, as well as support from USGS
518 Climate and Land Use Change Mission Area.

519 **List of Figures**

520	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].	4
521			
522			
523	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.	7
524			
525	3	Baseline cropland mask of Africa based on 10 pathfinding studies. Aggregated 250m cropland mask derived from 10 previous studies in Africa, including cropland fallow areas (Table 1).	8
526			
527	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	10
528			
529			
530			
531	5	Quantitative Spectral matching (QSM) of a generic class with an ideal spectral signatures.	11
532	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.	12
533			
534			
535			
536			
537	7	RCL2014 Spatial distribution. Reference cropland layers of Africa for the year 2014 (RCL2014). Three RCL2014 products:(a) Cropland <i>versus</i> non-Cropland Layer, (b) Irrigated <i>versus</i> Rainfed Layer, and (c) Crop intensities Layer.	14
538			
539			
540	8	RCL2014 seasonal cropland layers. Reference cropland product of the year 2014 (RCL2014) for Africa at 250m 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.	15
541			
542			
543			
544			
545	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, 2010).	17
546			
547	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.	19
548			
549			
550			
551			
552	11	ACL2003 to ACL2014 derived cropland areas <i>versus</i> cropland fallow areas.	21
553	12	ACMA derived croplands <i>versus</i> cropland fallows for a drought year (2005), normal year (2008), and wet year (2006) in AEZ3. The figure shows spatial distribution of croplands <i>versus</i> cropland fallows (left), mean MODIS 250-m NDVI during the three-year (top right) and precipitation (bottom right).	22
554			
555			
556	13	ACMA derived croplands <i>versus</i> cropland fallows for a drought year (2005), normal year (2008), and wet year (2006) in AEZ5. The figure shows spatial distribution of croplands <i>versus</i> cropland fallows (left), mean MODIS 250-m NDVI during the three-year (top right) and precipitation (bottom right).	23
557			
558			

559 **List of Tables**

560	1	Datasets used in creating 250m cropland mask of Africa in terms of their reference, data source, resolution, and time interval	5
561			
562	2	Description of crops mapped in global cropland product for Africa @ 250-m (GFSAD250).	9
563	3	RCL2014 overall accuracies for croplands <i>versus</i> non-croplands (product 1). Overall accuracies of RCL2014 product 1 (croplands <i>versus</i> 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 <i>versus</i> non-croplands (product 1) produced based on MODIS 250 m every 16 day NDVI data, ground data, and spectral matching techniques.	16
564			
565			
566			
567			
568	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.	16
569			
570			
571	5	RCL2014 overall accuracies for irrigated <i>versus</i> rainfed (product 2). Overall accuracies of irrigated <i>versus</i> rainfed RCL2014 product 2 (rainfed croplands <i>versus</i> 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 <i>versus</i> rainfed croplands (product 2) produced based on MODIS 250 m every 16 day NDVI data, ground data, and spectral matching techniques.	18
572			
573			
574			
575			
576	6	RCL2014 cropland area seasonal and total statistics of Africa for the year 2014 using MODIS 250m 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.	20
577			
578			
579			
580			
581	7	Similarity matrix between ACMA derived cropland product for the year 2014 (ACL2014) with reference cropland layer (RCL2014).	21
582			

583 **Reference**

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

588 Bégué, A., Vintrou, E., Saad, A., Hiernaux, P., 2014. Differences between cropland and rangeland MODIS phenology
589 (start-of-season) in Mali. *International Journal of Applied Earth Observation and Geoinformation* 31, 167.

590 Biradar, C.M., Thenkabail, P.S., Noojipady, P., Li, Y., Dheeravath, V., Turrall, H., Velpuri, M., Gumma, M.K., Gangalakunta,
591 O.R.P., Cai, X.L., Xiao, X., Schull, M.A., Alankara, R.D., Gunasinghe, S., Mohideen, S., 2009. A global map
592 of rainfed cropland areas (GMRCA) at the end of last millennium using remote sensing. *International Journal of
593 Applied Earth Observation and Geoinformation* 11, 114–129.

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

596 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,
597 J., 2015. Global land cover mapping at 30m resolution: A POK-based operational approach. *ISPRS Journal of
598 Photogrammetry and Remote Sensing* 103, 7.

599 Clover, J., 2010. FOOD SECURITY IN SUB-SAHARAN AFRICA. *African Security Studies* 12, 5–15.

600 Congalton, R., Gu, J., Thenkabail, P., Yadav, K., Ozdogan, M., 2014. Global Land Cover Mapping: A Review and Uncer-

601 tainty Analysis. *Remote Sensing* 6, 12070.

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

605 DeFries, R., 2000. Multiple Criteria for Evaluating Machine Learning Algorithms for Land Cover Classification from
606 Satellite Data. *Remote Sensing of Environment* 74, 503.

607 Deng, C., Wu, C., 2013. The use of single-date MODIS imagery for estimating large-scale urban impervious surface fraction
608 with spectral mixture analysis and machine learning techniques. *ISPRS Journal of Photogrammetry and Remote*
609 *Sensing* 86, 100.

610 Dheeravath, V., Thenkabail, P.S., Thenkabail, P.S., Noojipady, P., Chandrakantha, G., Reddy, G.P.O., Gumma, M.K., Biradar,
611 C.M., Velpuri, M., Gumma, M.K., 2010. Irrigated areas of India derived using MODIS 500 m time series for the
612 years 2001-2003. *ISPRS Journal of Photogrammetry and Remote Sensing* 65, 42.

613 Ding, Y., Ding, Y., Zhao, K., Zhao, K., Zheng, X., Zheng, X., Jiang, T., Jiang, T., 2014. Temporal dynamics of spatial
614 heterogeneity over cropland quantified by time-series NDVI, near infrared and red reflectance of Landsat 8 OLI
615 imagery. *International Journal of Applied Earth Observation and Geoinformation* 30, 139.

616 Dong, J., Xiao, X., Kou, W., Qin, Y., Zhang, G., Li, L., Jin, C., Zhou, Y., Wang, J., Biradar, C., Liu, J., Moore, B.,
617 2015. Tracking the dynamics of paddy rice planting area in 1986-2010 through time series Landsat images and
618 phenology-based algorithms. *Remote Sensing of Environment* 160, 99.

619 Dorward, A., Chirwa, E., 2010. A review of methods for estimating yield and production impacts.

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

623 Egorov, A.V., Hansen, M.C., Roy, D.P., Kommareddy, A., Potapov, P.V., 2015. Image interpretation-guided supervised
624 classification using nested segmentation. *Remote Sensing of Environment* 165, 135–147.

625 Esch, T., Metz, A., Marconcini, M., Keil, M., 2014. Combined use of multi-seasonal high and medium resolution satellite
626 imagery for parcel-related mapping of cropland and grassland. *International Journal of Applied Earth Observation*
627 *and Geoinformation* 28, 230–237.

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

630 Foody, G.M., 2010. Assessing the accuracy of land cover change with imperfect ground reference data. *Remote Sensing of*
631 *Environment* 114, 2271–2285.

632 Friedl, M.A., Brodley, C.E., 1997. Decision tree classification of land cover from remotely sensed data. *Remote Sensing of*
633 *Environment* 61, 399–409.

634 Friedl, M.A., McIver, D.K., Hodges, J., Zhang, X.Y., 2002. Global land cover mapping from MODIS: algorithms and early
635 results. *Remote Sensing of Environment* 287–302.

636 Fritz, S., McCallum, I., Schill, C., Perger, C., Grillmayer, R., Achard, F., Kraxner, F., Obersteiner, M., 2009. Geo-Wiki.Org:
637 The Use of Crowdsourcing to Improve Global Land Cover. *Remote Sensing* 1, 345–354.

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

640 Fritz, S., See, L., McCallum, I., Schill, C., Obersteiner, M., Velde, M. van der, Boettcher, H., Havlík, P., Achard, F., 2011.
641 Highlighting continued uncertainty in global land cover maps for the user community. *Environmental Research*

- 642 Letters 6, 044005.
- 643 Fritz, S., You, L., Bun, A., See, L., McCallum, I., Schill, C., Perger, C., Liu, J., Hansen, M., Obersteiner, M., 2011. Cropland
644 for sub-Saharan Africa: A synergistic approach using five land cover data sets. *Geophysical Research Letters* 38.
- 645 Funk, C.C., Peterson, P.J., Landsfeld, M.F., Pedreros, D.H., Verdin, J.P., Rowland, J.D., Romero, B.E., Husak, G.J.,
646 Michaelsen, J.C., Verdin, A.P., others, 2014. A quasi-global precipitation time series for drought monitoring. *US*
647 *Geological Survey Data Series* 832.
- 648 Gerland, P., Raftery, A.E., ikova, H. ev, Li, N., Gu, D., Spoorenberg, T., Alkema, L., Fosdick, B.K., Chunn, J., Lalic, N.,
649 Bay, G., Buettner, T., Heilig, G.K., Wilmoth, J., 2014. World population stabilization unlikely this century. *Science*
650 346, 234–237.
- 651 Giri, C., Zhu, Z., Reed, B., 2005. A comparative analysis of the Global Land Cover 2000 and MODIS land cover data sets.
652 *Remote Sensing of Environment* 94, 123–132.
- 653 Gumma, M.K., Gumma, M.K., Thenkabail, P.S., Maunahan, A., Islam, S., Nelson, A., 2014. Mapping seasonal rice cropland
654 extent and area in the high cropping intensity environment of Bangladesh using MODIS 500m data for the year
655 2010. *ISPRS Journal of Photogrammetry and Remote Sensing* 91, 98.
- 656 Haack, B., Mahabir, R., Kerkering, J., 2014. Remote sensing-derived national land cover land use maps: a comparison for
657 Malawi. *Geocarto International* 30, 270–292.
- 658 Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S.V., Goetz, S.J.,
659 Loveland, T.R., Kommareddy, A., Egorov, A., Chini, L., Justice, C.O., Townshend, J.R.G., 2013. High-Resolution
660 Global Maps of 21st-Century Forest Cover Change. *Science* 342, 850–853.
- 661 Hansen, M.C., Reed, B., 2010. A comparison of the IGBP DISCover and University of Maryland 1 km global land cover
662 products. *International Journal of Remote Sensing* 21, 1365–1373.
- 663 Helmholz, P., Rottensteiner, F., Heipke, C., 2014. Semi-automatic verification of cropland and grassland using very high
664 resolution mono-temporal satellite images. *ISPRS Journal of Photogrammetry and Remote Sensing* 97, 204–218.
- 665 Hentze, K., Thonfeld, F., Menz, G., 2016. Evaluating Crop Area Mapping from MODIS Time-Series as an Assessment Tool
666 for Zimbabwes Fast Track Land Reform Programme. *PLoS ONE* 11, e0156630.
- 667 Herold, M., Mayaux, P., Woodcock, C.E., Baccini, A., Schmullius, C., 2008. Some challenges in global land cover
668 mapping: An assessment of agreement and accuracy in existing 1 km datasets. *Remote Sensing of Environment*
669 112, 2538–2556.
- 670 Herold, M., Woodcock, C.E., Di Gregorio, A., Mayaux, P., Belward, A.S., Latham, J., Schmullius, C.C., 2006. A joint
671 initiative for harmonization and validation of land cover datasets. *IEEE Transactions on Geoscience and Remote*
672 *Sensing* 44, 1719–1727.
- 673 Jamali, S., Seaquist, J., Eklundh, L., Ardö, J., 2014. Automated mapping of vegetation trends with polynomials using NDVI
674 imagery over the Sahel. *Remote Sensing of Environment* 141, 79.
- 675 Jayne, T.S., Chamberlin, J., Headey, D.D., 2014. Land pressures, the evolution of farming systems, and development
676 strategies in Africa: A synthesis. *Food Policy* 48, 1–17.
- 677 Jeganathan, C., Dash, J., Atkinson, P.M., 2014. Remotely sensed trends in the phenology of northern high latitude terrestrial
678 vegetation, controlling for land cover change and vegetation type. *Remote Sensing of Environment* 143, 154.
- 679 Kidane, Y., Stahlmann, R., Beierkuhnlein, C., 2012. Vegetation dynamics, and land use and land cover change in the Bale
680 Mountains, Ethiopia. *Environmental Monitoring and Assessment* 184, 7473–7489.
- 681 Kruger, A.C., 2006. Observed trends in daily precipitation indices in South Africa: 19102004. *International Journal of*
682 *Climatology* 26, 2275–2285.
- 683 Kummu, M., De Moel, H., Porkka, M., Siebert, S., Varis, O., Ward, P.J., 2012. Lost food, wasted resources: Global food

- 684 supply chain losses and their impacts on freshwater, cropland, and fertiliser use. *Science of the Total Environment*
685 438, 477.
- 686 Lambert, M.-J., Waldner, F., Defourny, P., 2016. Cropland Mapping over Sahelian and Sudanian Agrosystems: A Knowledge-
687 Based Approach Using PROBA-V Time Series at 100-m. *Remote Sensing* 8, 232.
- 688 Lary, D.J., Alavi, A.H., Walker, A.L., Gandomi, A.H., 2016. Machine learning in geosciences and remote sensing.
689 *Geoscience Frontiers* 7, 3.
- 690 Latham, J., Cumani, R., Rosati, I., Bloise, M., 2014. Global land cover share (GLC-SHARE) database beta-release version
691 1.0-2014. FAO: Rome.
- 692 Leroux, L., Jolivot, A., Bégué, A., Seen, D., Zoungrana, B., 2014. How Reliable is the MODIS Land Cover Product for
693 Crop Mapping Sub-Saharan Agricultural Landscapes? *Remote Sensing* 6, 8541–8564.
- 694 Li, J., Liao, W.-k., Choudhary, A., Ross, R., Thakur, R., Gropp, W., Latham, R., Siegel, A., Gallagher, B., Zingale, M., 2003.
695 Parallel netCDF: A high-performance scientific I/O interface, in: *Supercomputing, 2003 Acm/Ieee Conference*.
696 IEEE, pp. 39–39.
- 697 Lobell, D.B., Asner, G.P., 2004. Cropland distributions from temporal unmixing of MODIS data. *Remote Sensing of*
698 *Environment* 93, 412–422.
- 699 Lobell, D.B., Burke, M.B., Tebaldi, C., Mastrandrea, M.D., Falcon, W.P., Naylor, R.L., 2008. Prioritizing Climate Change
700 Adaptation Needs for Food Security in 2030. *Science* 319, 607.
- 701 McCallum, I., Obersteiner, M., Nilsson, S., Shvidenko, A., 2006. A spatial comparison of four satellite derived 1km global
702 land cover datasets. *International Journal of Applied Earth Observation and Geoinformation* 8, 246–255.
- 703 Messerli, P., Messerli, P., 2009. Use of Sensitivity Analysis to Evaluate Key Factors for Improving Slash-and-Burn
704 Cultivation Systems on the Eastern Escarpment of Madagascar. *dx.doi.org* 20, 32–41.
- 705 Motha, R.P., Leduc, S.K., Steyaert, L.T., Sakamoto, C.M., Strommen, N.D., Motha, R.P., Leduc, S.K., Steyaert, L.T.,
706 Sakamoto, C.M., Strommen, N.D., 1980. Precipitation Patterns in West Africa. [http://dx.doi.org/10.1175/1520-0493\(1980\)108<1567:PPIWA>2.0.CO;2](http://dx.doi.org/10.1175/1520-0493(1980)108<1567:PPIWA>2.0.CO;2) 108, 1567–1578.
- 708 Mountrakis, G., Im, J., Ogole, C., 2011. Support vector machines in remote sensing: A review. *ISPRS Journal of*
709 *Photogrammetry and Remote Sensing* 66, 247.
- 710 Müller, H., Rufin, P., Griffiths, P., Barros Siqueira, A.J., Hostert, P., 2015. Mining dense Landsat time series for separating
711 cropland and pasture in a heterogeneous Brazilian savanna landscape. *Remote Sensing of Environment* 156,
712 490–499.
- 713 Nemani, R., Votava, P., Michaelis, A., Melton, F., Milesi, C., 2011. Collaborative Supercomputing for Global Change
714 Science. *Eos, Transactions American Geophysical Union* 92, 109–110.
- 715 Ozdogan, M., Gutman, G., 2008. A new methodology to map irrigated areas using multi-temporal MODIS and ancillary
716 data: An application example in the continental US. *Remote Sensing of Environment* 112, 3520–3537.
- 717 Pan, Z., Huang, J., Zhou, Q., Wang, L., Cheng, Y., Zhang, H., Blackburn, G.A., Yan, J., Liu, J., 2015. Mapping crop
718 phenology using NDVI time-series derived from HJ-1 A/B data. *International Journal of Applied Earth Observation*
719 *and Geoinformation* 34, 188–197.
- 720 Pantazi, X.E., Moshou, D., Alexandridis, T., Mouazen, A.M., Whetton, R.L., 2016. Wheat yield prediction using machine
721 learning and advanced sensing techniques. *Computers and Electronics in Agriculture* 121, 57.
- 722 Peña-Arancibia, J.L., Mainuddin, M., Kirby, J.M., Chiew, F.H.S., McVicar, T.R., Vaze, J., 2016. Assessing irrigated
723 agriculture’s surface water and groundwater consumption by combining satellite remote sensing and hydrologic
724 modelling. *The Science of the total environment* 542, 372–382.
- 725 Pittman, K., Hansen, M.C., Becker-Reshef, I., Potapov, P.V., Justice, C.O., 2010. Estimating Global Cropland Extent with

- 726 Multi-year MODIS Data. *Remote Sensing* 2, 1844–1863.
- 727 Portmann, F.T., Siebert, S., Döll, P., 2010. MIRCA2000 Global monthly irrigated and rainfed crop areas around the year
728 2000: A new high-resolution data set for agricultural and hydrological modeling. *Global Biogeochemical Cycles*
729 24, 1–24.
- 730 Qiu, B., Zhong, M., Tang, Z., Wang, C., 2014. A new methodology to map double-cropping croplands based on continuous
731 wavelet transform. *International Journal of Applied Earth Observation and Geoinformation* 26, 97–104.
- 732 Radoux, J., Lamarche, C., Van Bogaert, E., Bontemps, S., Brockmann, C., Defourny, P., 2014. Automated Training Sample
733 Extraction for Global Land Cover Mapping. *Remote Sensing* 6, 3965–3987.
- 734 Rembold, F., Carnicelli, S., Nori, M., Ferrari, G.A., 2000. Use of aerial photographs, Landsat TM imagery and multidisci-
735 plinary field survey for land-cover change analysis in the lakes region (Ethiopia). *International Journal of Applied*
736 *Earth Observation and Geoinformation* 2, 181–189.
- 737 Salmon, J.M., Friedl, M.A., Froking, S., Wisser, D., Douglas, E.M., 2015. Global rain-fed, irrigated, and paddy croplands:
738 A new high resolution map derived from remote sensing, crop inventories and climate data. *International Journal*
739 *of Applied Earth Observation and Geoinformation* 38, 321–334.
- 740 See, L., Comber, A., Salk, C., Fritz, S., Velde, M. van der, Perger, C., Schill, C., McCallum, I., Kraxner, F., Obersteiner, M.,
741 2013. Comparing the quality of crowdsourced data contributed by expert and non-experts. *PLoS ONE* 8, e69958.
- 742 Shalaby, A., Tateishi, R., 2007. Remote sensing and GIS for mapping and monitoring land cover and land-use changes in
743 the Northwestern coastal zone of Egypt. *Applied Geography* 27, 28–41.
- 744 Shao, Y., Lunetta, R.S., 2012. Comparison of support vector machine, neural network, and CART algorithms for the
745 land-cover classification using limited training data points. *ISPRS Journal of Photogrammetry and Remote Sensing*
746 70, 78.
- 747 Sharma, R., Ghosh, A., Joshi, P.K., 2013. Decision tree approach for classification of remotely sensed satellite data using
748 open source support. *Journal of Earth System Science* 122, 1237.
- 749 Tateishi, R., Hoan, N.T., Kobayashi, T., 2014. Production of Global Land Cover Data GLCNMO2008. *Journal of Geography*
750 *and Geology* 6, 99.
- 751 Tatsumi, K., Yamashiki, Y., Canales Torres, M.A., Taibe, C.L.R., 2015. Crop classification of upland fields using Random
752 forest of time-series Landsat 7 ETM+ data. *Computers and Electronics in Agriculture* 115, 171–179.
- 753 Teluguntla, P., Ryu, D., George, B., Walker, J.P., Malano, H.M., 2015. Mapping flooded rice paddies using time series of
754 modis imagery in the krishna river basin, india. *Remote Sensing* 7, 8858–8882.
- 755 Teluguntla, P., Thenkabail, P., Xiong, J., Gumma, M.K., Giri, C., Milesi, C., Ozdogan, M., Congalton, R., Yadav, K., 2015.
756 CHAPTER 6 - Global Food Security Support Analysis Data at Nominal 1 km (GFSAD1km) Derived from Remote
757 Sensing in Support of Food Security in the Twenty-First Century: Current Achievements and Future Possibilities,
758 in: Thenkabail, P.S. (Ed.), CRC Press, Boca Raton, London, New York., pp. 131–160.
- 759 Teluguntla, P., Thenkabail, P.S., Xiong, J., Gumma, M.K., Congalton, R.G., Oliphant, A., Poehnelt, J., Yadav, K., Rao,
760 M.N., Massey, R., 2016. Spectral matching techniques (smts) and automated cropland classification algorithms
761 (accas) for mapping croplands of australia using modis 250-m time-series (2000-2015) data. *International Journal*
762 *of Digital Earth* 1–36.
- 763 Thenkabail, P., Gangadhara Rao, P., Biggs, T., 2007. Spectral matching techniques to determine historical land-use/land-
764 cover (LULC) and irrigated areas using time-series 0.1-degree AVHRR Pathfinder datasets. *Photogrammetric*
765 *Engineering & Remote Sensing* 73, 1029–1040.
- 766 Thenkabail, P.S., Wu, Z., 2012. An Automated Cropland Classification Algorithm (ACCA) for Tajikistan by Combining
767 Landsat, MODIS, and Secondary Data. *Remote Sensing* 4, 2890–2918.
- 768 Thenkabail, P.S., Biradar, C.M., Noojipady, P., Cai, X., Dheeravath, V., Li, Y., Velpuri, M., Gumma, M., Pandey, S., 2007.

- 769 Sub-pixel Area Calculation Methods for Estimating Irrigated Areas. *Sensors* 7, 2519.
- 770 Tsendbazar, N.E., Tsendbazar, N.E., Bruin, S. de, Herold, M., Herold, M., 2015. Assessing global land cover reference
771 datasets for different user communities. *ISPRS Journal of Photogrammetry and Remote Sensing* 103, 93.
- 772 Vancutsem, C., Marinho, E., Kayitakire, F., See, L., Fritz, S., 2012. Harmonizing and Combining Existing Land Cover/Land
773 Use Datasets for Cropland Area Monitoring at the African Continental Scale. *Remote Sensing* 5, 19–41.
- 774 Vintrou, E., Desbrosse, A., Bégué, A., Traoré, S., 2012. Crop area mapping in West Africa using landscape stratification of
775 MODIS time series and comparison with existing global land products. *International Journal of . . .* 14, 83–93.
- 776 Waldner, F., Canto, G.S., Defourny, P., 2015. Automated annual cropland mapping using knowledge-based temporal features.
777 *ISPRS Journal of Photogrammetry and Remote Sensing* 110, 1–13.
- 778 Waldner, F., De Abelleira, D., Verón, S.R., Zhang, M., Wu, B., Plotnikov, D., Bartalev, S., Lavreniuk, M., Skakun, S.,
779 Kussul, N., Le Maire, G., Dupuy, S., Jarvis, I., Defourny, P., 2016. Towards a set of agrosystem-specific cropland
780 mapping methods to address the global cropland diversity. *International Journal of Remote Sensing* 37, 3196–3231.
- 781 Wang, J., Zhao, Y., Yu, L., Li, C., Yu, Liu, D., Gong, P., 2015. Mapping global land cover in 2001 and 2010 with
782 spatial-temporal consistency at 250m resolution. *ISPRS Journal of Photogrammetry and Remote Sensing* 103, 38.
- 783 Were, K.O., Dick, Ø.B., Singh, B.R., 2013. Remotely sensing the spatial and temporal land cover changes in Eastern Mau
784 forest reserve and Lake Nakuru drainage basin, Kenya. *Applied Geography* 41, 75–86.
- 785 Wu, Z., Verdin, J.P., Thenkabail, P.S., Verdin, J.P., 2014. Automated Cropland Classification Algorithm (ACCA) for
786 California Using Multi-sensor Remote Sensing. *Photogrammetric Engineering and Remote Sensing* 80, 81.
- 787 Yan, L., Roy, D.P., 2014. Automated crop field extraction from multi-temporal Web Enabled Landsat Data. *Remote Sensing
788 of Environment* 144, 42.
- 789 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.,
790 Menarguez, M.A., Biradar, C., 2015. Mapping paddy rice planting areas through time series analysis of MODIS
791 land surface temperature and vegetation index data. *ISPRS Journal of Photogrammetry and Remote Sensing* 106,
792 157.
- 793 Zhang, J., Wu, G., Hu, X., Li, S., Hao, S., 2011. A Parallel K-Means Clustering Algorithm with MPI. *IEEE*.
- 794 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,
795 X., 2014. Towards a common validation sample set for global land-cover mapping. *International Journal of Remote
796 Sensing* 35, 4795–4814.
- 797 Zhong, L., Gong, P., Biging, G.S., 2014. Efficient corn and soybean mapping with temporal extendability: A multi-year
798 experiment using Landsat imagery. *Remote Sensing of Environment* 140, 1.
- 799 Zhou, Y., Xiao, X., Qin, Y., Dong, J., Zhang, G., Kou, W., Jin, C., Wang, J., Li, X., 2016. Mapping paddy rice planting area
800 in rice-wetland coexistent areas through analysis of Landsat 8 OLI and MODIS images. *International Journal of
801 Applied Earth Observation and Geoinformation* 46, 1–12.
- 802 Zucca, C., Wu, W., Dessena, L., Mulas, M., 2015. Assessing the Effectiveness of Land Restoration Interventions in Dry
803 Lands by Multitemporal Remote Sensing A Case Study in Ouled DLIM (Marrakech, Morocco). *Land Degradation
804 & Development* 26, 80–91.