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

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

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

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

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

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

ARTICLE INFO

Article history:
Received 11 July 2016
Received in revised form 9 December 2016
Accepted 24 January 2017
Available online 8 March 2017

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

ABSTRACT

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

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

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

http://dx.doi.org/10.1016/j.isprsjprs.2017.01.019

1. Introduction

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

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

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

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

Further, the current state-of-art using the above methods and approaches is mostly limited to producing cropland products for a given period, or for a growing season, or for a particular year. However, such a process over very large areas such as continent will always have limitations in availability of extensive collection of reference data. The biggest difficulty in cropland mapping is in the lack of algorithms that accurately reproduce cropland products year after year or season after season. So, more recently, there are efforts at producing cropland products by developing automated algorithms (Jamali et al., 2014; Waldner et al., 2015; Yan and Roy, 2014). Thenkabail et al. developed rule-based ensemble decision-tree Automated Cropland Classification algorithms (ACCA's) to produce cropland *versus* non-croplands across years for Australia, Tajikistan and California (Teluguntla et al., 2017; Thenkabail and Wu, 2012; Wu et al., 2014). Waldner et al. (2015) used a baseline map generated from five knowledge-based temporal features to train an automated support vector machines (SVM) classifier on selected areas in Argentina, Belgium, Ukraine, and