DCN
Version 1.0

September 2017 Version

NASA Making Earth System Data Records for Use in Research Environments (MEaSUREs) Global Food Security-support Analysis Data (GFSAD) 30-m Cropland Extent Product of South Asia, Afghanistan and Iran (GFSAD30SAAFGIRCE)

Algorithm Theoretical Basis Document (ATBD)

USGS EROS
Sioux Falls, South Dakota

Document History

Document Version	Publication Date	Description
1.0	September 2017	Original

Contents

I. Members of the team

This Global Food Security-support Analysis Data 30-m (GFSAD30) Cropland Extent-Product of South Asia, Afghanistan and Iran (**GFSAD30SAAFGIRCE**) was produced by the following team members. Their specific role is mentioned.

Dr. Murali Krishna Gumma, Senior Scientist at the International Crops Research Institute for the Semi-Arid Tropics (***ICRISAT***) led the GFSAD30SAAFGIRCE product generation effort. Dr. Gumma was instrumental in the designing, coding, computing, analyzing, and synthesis of the Landsat-8 derived nominal 30-m GFSAD30SAAFGIRCE

Dr. Prasad S. Thenkabail, Research Geographer, United States Geological Survey, is the Principal Investigator (PI) of the GFSAD30 project. Dr. Thenkabail was instrumental in developing the conceptual framework of the GFSAD30 project and the GFSAD30SAAFGIRCE product. He made significant contribution in writing the manuscripts, ATBDs, User documentations, and providing scientific guidance on the GFSAD30 project.

Dr. Pardhasaradhi Teluguntla, Research Scientist, Bay Area Environmental Research Institute (BAERI) at United States Geological Survey (USGS, shared his expertise in cloud computing and Random Forest algorithm implementation in Google Earth Engine (GEE) for GFSAD30 SAAFGIRCE 30-m cropland extent product generation. He was also instrumental in writing the manuscripts, ATBDs, and user documentations.

Mr. Adam Oliphant, Geographer**,** United States Geological Survey (USGS), shared his expertise in cloud computing and Random Forest algorithm implementation in GEE for GFSAD30 SAAFGIRCE 30-m cropland extent product generation.

Dr. Jun Xiong, Research Scientist, Bay Area Environmental Research Institute (BAERI) at United States Geological Survey (USGS), participated in the intellectual discussions and in provided inputs and insights on GFSAD30 SAAFGIRC**E** 30-m cropland extent product generation and shared his expertise in cloud computing.

Dr. Russell G. Congalton, Professor of Remote Sensing and GIS at the University of New Hampshire, led the independent accuracy assessment of the entire GFSAD30 project including GFSAD30 SAAFGIRCE 30-m cropland extent product of Australia, New Zealand and China.

Ms. Kamini Yadav, PhD student at the University of New Hampshire, made contributions to the independent accuracy assessment directed by Prof. Russell G. Congalton.

Ms. Corryn Smith, Student developer, helped in development of the croplands.org website.

II. Historical Context and Background Information

Monitoring global croplands (GCs) is imperative for ensuring sustainable water and food security for the people of the world in the Twenty-first Century. However, currently available cropland products suffer from major limitations such as: (a) Absence of precise spatial location of the cropped areas; (b) Coarse resolution nature of the map products with significant uncertainties in areas, locations, and detail; (c) Uncertainties in differentiating irrigated areas from rainfed areas; (d) Absence of crop types and cropping intensities; and (e) Absence of a dedicated web\data portal for the dissemination of cropland products. Therefore, the Global Food Security-support Analysis Data (GFSAD) project aimed to address these limitations by producing cropland maps at 30m resolution covering the globe, referred to as Global food security support-analysis data @ 30-m (GFSAD30) product.

This Algorithm Theoretical Basis Document (ATBD) provides a basis upon which the GFSAD30 cropland extent product was developed for the countries of South Asia, Afghanistan and Iran, (GFSAD30SAAFGIRCE, Table 1), produced using Landsat-8 and Landsat-7 time-series satellite sensor data. This document provides comprehensive details of the GFSAD30CSAAFGIRCE production scheme that includes remote sensing data, reference and validation data, approaches, methods, machine learning algorithms, product generation, accuracy assessments, and area calculations.

Table 1. GFSAD30CE Product basic information for South Asia, Afghanistan and Iran.

Product Name	Short Name	Spatial resolution	Temporal coverage
GFSAD30-m cropland Extent Product of South Asia, Afghanistan and Iran	GFSAD30SAAFGIR	30-m	Nominal 2015

Note: Nominal here means that the Landsat-8 16 day data used to produce the product is for two to three years (2013-2015), but the product is reported as nominal year 2015.

III. Rationale for Development of the Algorithms

Mapping the precise location of croplands enables the extent and area of agricultural lands to be more effectively captured, which is of great importance for managing food production systems and to study their inter-relationships with water, geo-political, socio-economic, health, environmental, and ecological issues (Thenkabail et al., 2010). Further, accurate development of all higher-level cropland products such as crop watering method (irrigated or rainfed), cropping intensities (e.g., single, double, or continuous cropping), crop type mapping, cropland fallows, as well as assessment of cropland productivity (i.e., productivity per unit of land), and crop water productivity (i.e., productivity per unit of water) are all highly dependent on availability of precise and accurate cropland extent maps. Uncertainties associated with cropland extent maps affect the quality of all higher-level cropland products reliant on an accurate cropland extent base map. However, precise and accurate cropland extent maps are currently nonexistent at the continental extent at a high spatial resolution (30-m or better). This lack of crop extent maps is particularly true for complex, small-holder dominant agricultural systems. By mapping croplands at a high spatial resolution at the continental scale, the GFSAD30project has resolved many of the shortcomings and uncertainties of other cropland mapping efforts.

The two most common methods for land-cover mapping over large areas using remote-sensing images are manual classification based on visual interpretation and digital per-pixel classification. The former approach delivers products of high quality, such as the European CORINE Land Cover maps (Büttner, 2014). Although the human capacity for interpreting images is remarkable, visual interpretation is subjective (Lillesand et al., 2014), time-consuming, and expensive. Digital per-pixel classification has been applied for land-cover mapping since the advent of remote sensing and is still widely used in operational programs, such as the 2005 North American Land Cover Database at 250-m spatial resolution (Latifovic, 2010). Pixel-based classifications such as maximum likelihood classifier (MLC), neural network classification (NN), decision trees, Random Forests (RF), and Support Vector Machines are powerful, and fast classifiers that help differentiate distinct patterns of landscape. Both supervised and unsupervised classification approaches are adopted in pixel-based classifiers. However, per-pixel classification includes several limitations. For example, the pixel's square shape is arbitrary in relation to patchy or continuous land features of interest, and there is significant spectral contamination among neighboring pixels. As a result, per-pixel classification often leads to noisy classification outputs, the well-known "salt-and-pepper" effect. There are other limitations of pixel-based classification methods: 1. they fail to capture the spatial information of high-resolution imagery such as from Landsat 30-m imagery, and 2. they often, classify the same field (e.g., a corn field) into different classes, as a result of within-field variability. This may often result in a field with a single crop (e.g., corn) classified as different crops.

In this study, we the used supervised pixel-based classifier Random Forest (RF), which has been widely used in agricultural cropland studies over the years (Myint et al., 2011) and which is considered powerful and an ideal machine learning algorithm (Tian et al., 2016, Shi and Yang, 2015, Huang et al., 2010). A description of how to classify cropland extent of South Asia, Afghanistan and Iran is provided in section 2.3 and its sub-sections.

This document describes, in detail, the development of the 30-m Cropland Extent Product of South Asia, Afghanistan and Iran (GFSAD30SAAFGIRCE). The approach involves the use of a supervised Random Forest (RF) algorithm to retrieve crop extent results from pixel-based classification (see overview of the methodology in Figure 1).

IV. Algorithm Description

An overview of the algorithm description is provided in Figure 1. The methodology used in this project (Figure 1) is briefly described in this paragraph to provide an overview of methods presented in detail in subsequent sections of this document. The process (Figure 1) involved combining year 2013-2016 16-day time-series Landsat-8 30-m data. The process included several well designed steps (Figure 1). First, the data were pre-processed by cloud mask and gap filling on GEE. Second, median value composites were created for three seasons/periods based on cloud-free or near-cloud-free wall-to-wall coverage. Such a seasonal mosaic aided in achieving cloud free clear images of the study area. Each composite mosaic contained 10 bands as listed in Figure 1. Third, reference data were generated throughout the study area to train the RF algorithms. There are 7976 reference samples for this purpose. Fourth, the result of the pixel-based RF algorithm was to obtain the cropland extent product for South Asia, Afghanistan and Iran. Fifth, the cropland product of South Asia, Afghanistan and Iran was evaluated for accuracy us-

ing 1200 test samples. The process was iterated until adequate accuracies were attained. Accuracy assessments were performed by Dr. Russell Congalton and his PhD student, Kamini Yadav, independent of the production team. In this process, the validation data were only available to the accuracy assessment team and were hidden from the production team. As a result, there was completely independent accuracy assessment. Finally, the GFSAD30SAAFGIRCE product was made available on croplands.org.

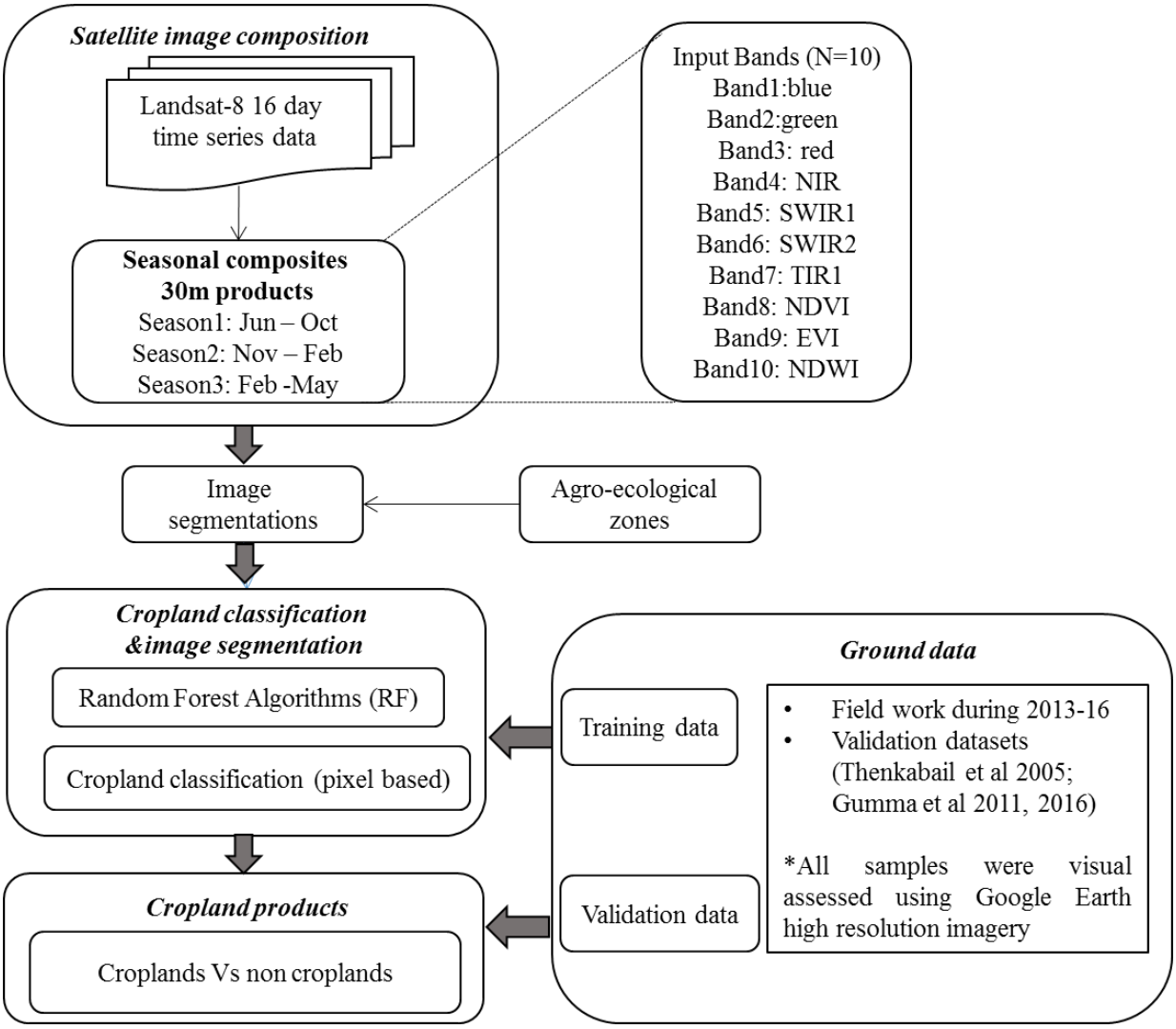

Figure 1. Flowchart of mapping methods for Landsat-8 derived cropland extent-product of South Asia, Afghanistan, and Iran for the nominal year 2015.

a. Input data

i. Region Definition

The study was conducted for all the countries of South Asia region, Afghanistan, and Iran (see Figure 2 and Figure 5). The country boundaries were determined by the Global Administrative Unit Layers (GAUL) of United Nations (http://www.fao.org/geonetwork/srv/en/metadata.show?id=12691&currTab=simple).

ii. Reference Croplands Samples

Reference data are required for both training the machine learning algorithms (see section 2.3) as well as for validating the final products. First, we conducted extensive field survey between 2013- 2016, during the crop-growing season for major crops in study regions.

More than 2800 ground samples were collected from our study Area following the specific guidelines on collecting ground reference data (Congalton, 2015). The sampling sites included various crop types: such as Cereal crops (Rice, Wheat, Maize, and Sorghum), Millets, Legumes (Pigeon pea, Chickpea, Black-gram, Green-gram, Lentils, Peas, and Beans), Oilseeds (Groundnut, sunflower, and Cotton), Vegetables, Continuous crops (Sugarcane, Orchard crops, Plantations), Fodder crops and some fallow lands. The field survey gathered more than 2800 ground samples including: (1) Location of samples (GPS position, location name, date of collection); crop properties (2) Croplands versus non-croplands; (3) irrigated or rainfed; (4) Crop intensity (single, double, triple, continuous cropping in 12 months); (5) Crop type (major crop types mentioned above, others); and (6) Digital photographs of each sample.

The ground data samples were collected from three main sources.

First, field surveys (or ground data) were collected during 2013 to 2016. The field-surveyed data were divided into two independent datasets with 60:40 split. The first set was used for training machine-learning algorithms (e.g., Random Forest) and testing the product. The second set was kept aside and was used for independent accuracy assessment. Also, we obtained reference-training data from the following reliable sources in addition to our own field data collections.

Second, random samples were obtained by interpreting sub-meter to 5-meter very high spatial resolution imagery (VHRI) data throughout South Asia, Afghanistan and Iran available to US Government entities through the sub-meter to 5-m imagery obtained from the National Geospatial Agency (NGA). For this, we collected 5176 reference and 1500 validation samples.

Third, reference data were obtained from other reliable sources such as Central Research Institute for Dryland Agriculture (CRIDA) a national Institute in India, Department of Agriculture Extension, Bangladesh in Bangladesh. The reference training data were used to “train” the Random Forest algorithm to separate croplands from non-croplands. This required us to keep adding training samples until optimal classification results were obtained (see section 2.3 and its sub-sections). A total of 7976 (crop= 3314, no crop= 4662) representative samples were used to “train” and separate croplands from non-croplands in South Asia, Afghanistan and Iran (see Figure 2 showing the distribution of these samples) (Table 2).

The whole set of Reference data including primary and secondary data were made available, at the following website: https://croplands.org/app/data/search

Table 2. Number of reference samples used for training the Random Forest (RF) machine-learning algorithm and number of validation samples used for independent accuracy assessment.

Zone#	Class	Training samples	Validation samples
	Crop	1889	59
Zone1	No Crop	1485	191
	Total	**3374**	**250**
	Crop	460	12
Zone2	No Crop	2008	238
	Total	**2468**	**500**
	Crop	144	151
Zone3	No Crop	87	99
	Total	231	**250**
	Crop	392	164
Zone4	No Crop	507	85
	Total	**899**	**250**
	Crop	346	140
Zone5	No Crop	457	109
	Total	**803**	**250**
	Crop	83	31
Zone6	No Crop	118	219
	Total	**201**	**250**
	Crop	3314	557
Total	No Crop	4662	941
6 zones	**Total**	**7976**	**1500**

Note: The number of training and validation samples depended on the results. When optimal results obtained, we stopped adding further samples. The process requires starting with a certain sample number initially and progressively increasing sample number until optimal accuracies are reached.

iii. Image Stratification

The cropland versus non-cropland classification was carried out using the Random Forest (RF) machine-learning algorithm by stratifying the study area into refined agro-ecological zones (AEZs) (Figure 2). The AEZs were developed by the United Nation's Food and Agricultural Organization (UN FAO). However, this results in too many zones (which is not necessary given many zones have only very small proportion of crops). Therefore, we combined some of these zones into broader refined AEZs (RAEZs) based on the convenience, and speed of applying the RF algorithm. This resulted in six broad RAEZs across South Asia, Afghanistan and Iran (Figure 2). RF algorithm were trained for separating croplands from non-croplands in each of these RAEZs (Figure 2) using the reference training data falling within these zones. Working within each RAEZ also helped in data management and classification speed.

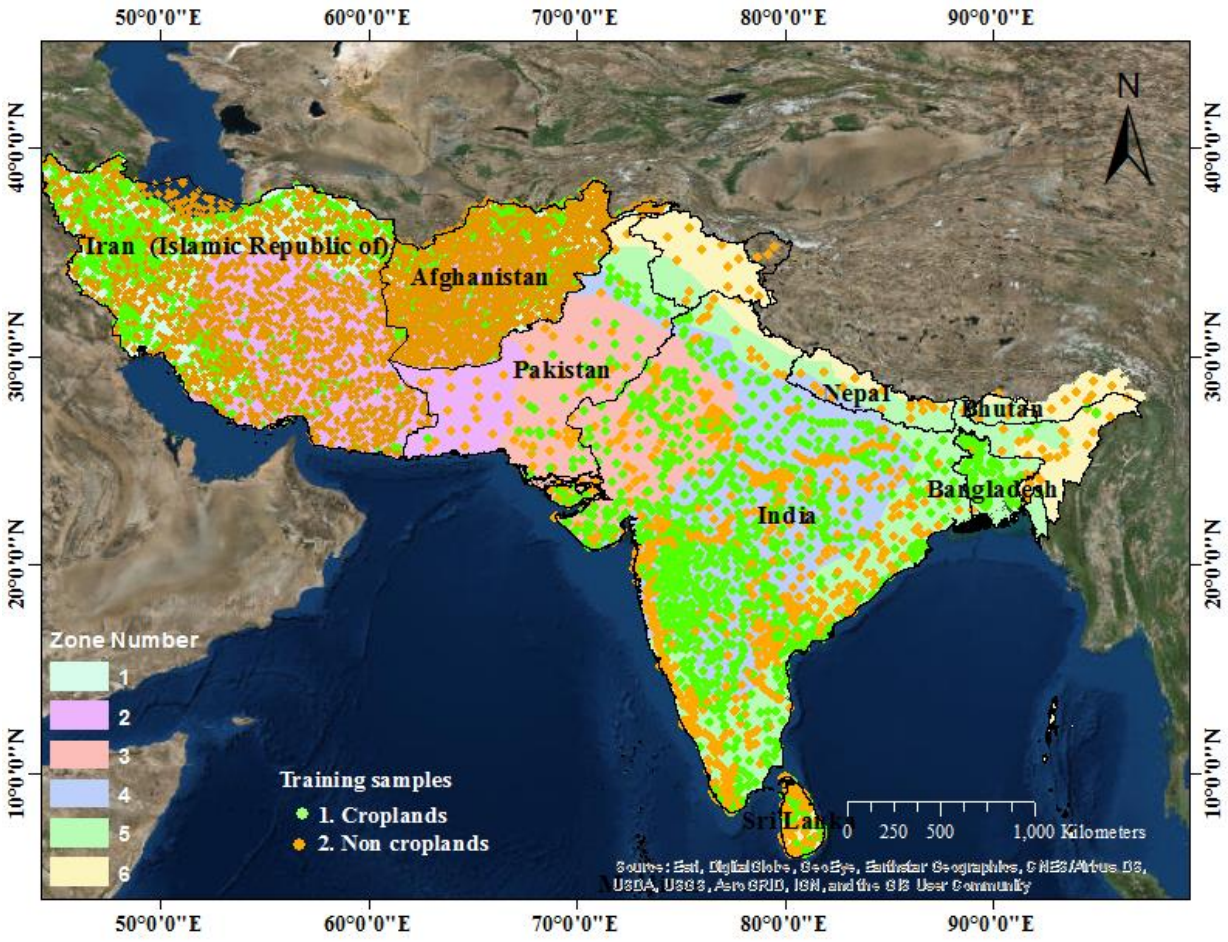

Figure 2. Stratification of the study area into distinct and broad refined agro-ecological zones (RAEZs). The figure also shows the distribution of the reference training data used in the Random Forest (RF) machine-learning algorithm. The Random Forest (RF) pixel-based supervised machine learning algorithm used in this study was "trained" using reference training data falling within each of these zones to separate croplands from non-croplands.

iv. Satellite Imagery: Landsat-8

In order to cover crop dynamics in different periods/seasons, Landsat-8 OLI (Roy et al., 2014) satellite data have been used for South Asia, Afghanistan and Iran. There is a 16-day revisit time per Landsat-8 OLI 30-m data. It is difficult to get continuous 16-day cloud free time-series data for wall-to-wall coverage for any part of the region. To overcome this limitation and to ensure cloud-free or near-cloud-free wall-to-wall coverage, bi/tri-monthly seasonal composites, depending on the cloudiness of the countries\regions, were composed (e.g., Figures 3). Finally, 30-m mega-file data-cube (MFDC) were created as per the following steps leading to a 32-band MFDC (Figure 3) for South Asia, Afghanistan and Iran from three seasons/periods. A systematic detail of the MFDC composition is described below.

The goal of the time-composites was to achieve cloud-free or near cloud-free wall-to-wall composites over the entire study area (e.g., Figures 3). This we wanted to achieve, using as many time-periods as possible as to get temporal stacks that can monitor phenology. However, the time-periods are decided by the ability to achieve cloud-free or near cloud-free images over a time-period. Based on crop seasons in the study region, we were able to achieve the cloud-free or near cloud-free images at much shorter time-periods leading to 3 seasons/periods (period1 (kharif season / monsoon season) Julian days 151-300; period 2 (rabi season / winter season): Julian days 301-365,1-60; period 3 (summer season): Julian days 61-150)(Figure 3).

The process involved gathering all the Landsat-8 16-day images over South Asia, Afghanistan and Iran.(Figure 3), available for each time-period/season (e.g., period1(kharif season) Julian days 151-300), and compositing each of the 10 bands by taking median value of each pixel of each band. These composites are called median value composites for each period for each band. The ten bands used in this study were (Figure 3): blue (0.45-0.51μm), green (0.53-0.59μm), red 0.63-0.69μm), NIR (0.85-0.89μm), SWIR1 (1.55 1.65μm), SWIR2 (2.1-2.3 μm), and TIR1 (10.60-11.19μm) bands along with Normalized Difference in Vegetation Index (NDVI), Enhanced Vegetation Index (EVI) and Normalized difference in water Index(NDWI). Thereby, for South Asia, Afghanistan and Iran, ten median value bands composed over 3 time-periods/ seasons resulted in a 30 bands. Additionally elevation and slope also added finally resulted in 32-band MFDC (Figure 3). The band stack, and time-periods leading to MFDC are shown in Table 3 as well as in Figures 3. All compositions were performed on the GEE cloud-based geospatial platform for planetary-scale data analysis (Gorelick et al,, 2017). Landsat top of atmosphere (TOA) products were used instead of surface reflectance (SR) due to the limited temporal availability of Landsat-8 surface reflectance imagery on GEE.

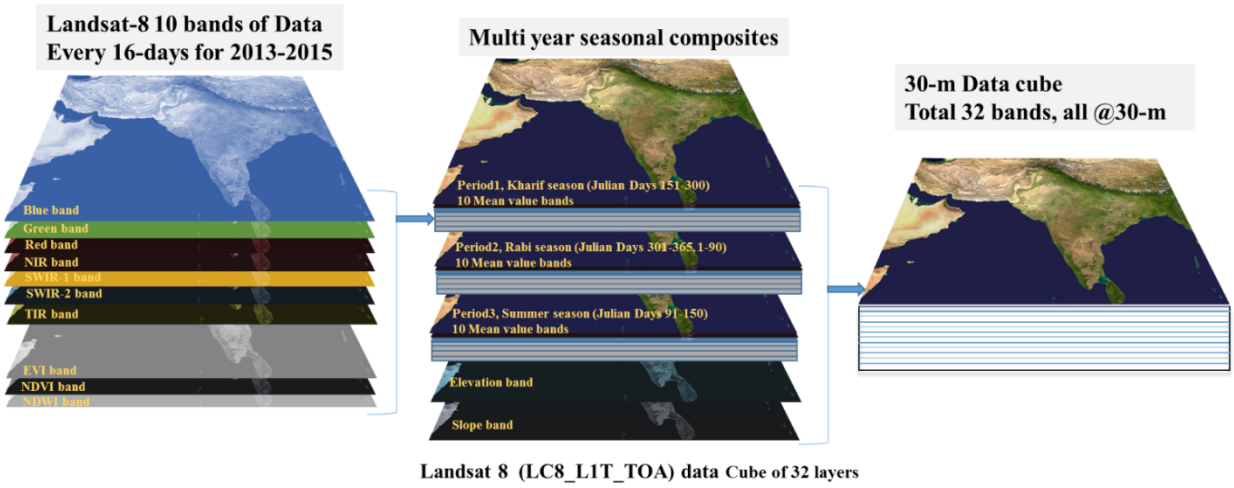

Figure 3. 30-m Data-cube for the South Asia, Afghanistan and Iran regions composited for 3 time-periods/ seasons using 2013-2016 Landsat-8 data. For each period (e.g., period 1 (kharif season: Julian days 151-300), ten bands (blue, green, red, NIR, SWIR1, SWIR2, TIR1, NDVI, EVI and NDWI Landsat-8) were composited, taking median value of a given pixel over the period 1. From 3 periods, and elevation and slope, so there was a 32-band mega file data cube (MFDC).

Table 3. The process of mega file data cube (MFDC) composition for the study areas based on median value composition of 8 Landsat-8 bands over 2013-2016 for three time-periods/seasons.

SN#	**Region/ Country**	**Landsat products/ image Series**	**Years of data**	**Time composited***	**Bands used****	**Mega-file Data Cube (MFDC)**
#	**name**	**satellite, sensor**	**Years**	**Julian days over each composited period**	**# of bands for each period**	**Total # of bands in MFDC**
1	South Asia, Afghanistan and Iran	Landsat-8 Data source: GEE	2014 - 2016	C1:151-300 C2:301-365, and 1-60 C3:61-150	Blue, green, Red, NIR , SWIR-1, SWIR-2, Temperature, NDVI, EVI and NDWI	32

***** C1:151-300 = composite 1 over Julian dates 151 to 300. Given Landsat-8 is acquired over every 16 days,
There will be ~9 to 10 images in first season.
Then each band (e.g., blue) is derived using median value from these 10 images. Similarly for all bands.
Similarly for other periods/seasons

** NIR - near-infrared, SWIR = short-wave infrared, NDVI = normalized difference vegetation index,
EVI= enhanced vegetation index and NDWI = normalized difference water index.

b. Theoretical Description

i. Definition of Croplands

For all products within GFSAD30 cropland extent map, cropland extent was defined as, "lands cultivated with plants harvested for food, feed, and fiber, including both seasonal crops (e.g., wheat, rice, corn, soybeans, cotton) and continuous plantations (e.g., coffee, tea, rubber, cocoa, and oil palms). Cropland fallows are lands uncultivated during a season or a year but are farmlands and are equipped for cultivation, including plantations (e.g., orchards, vineyards, coffee, tea, and rubber)" (Teluguntla et al., 2015). Cropland extent also includes areas equipped for cropping but may not be cropped in a particular season or year. These are cropland fallows. So cropland extent includes all planted crops plus cropland fallows. Non-croplands include all other land cover classes other than croplands and cropland fallows.

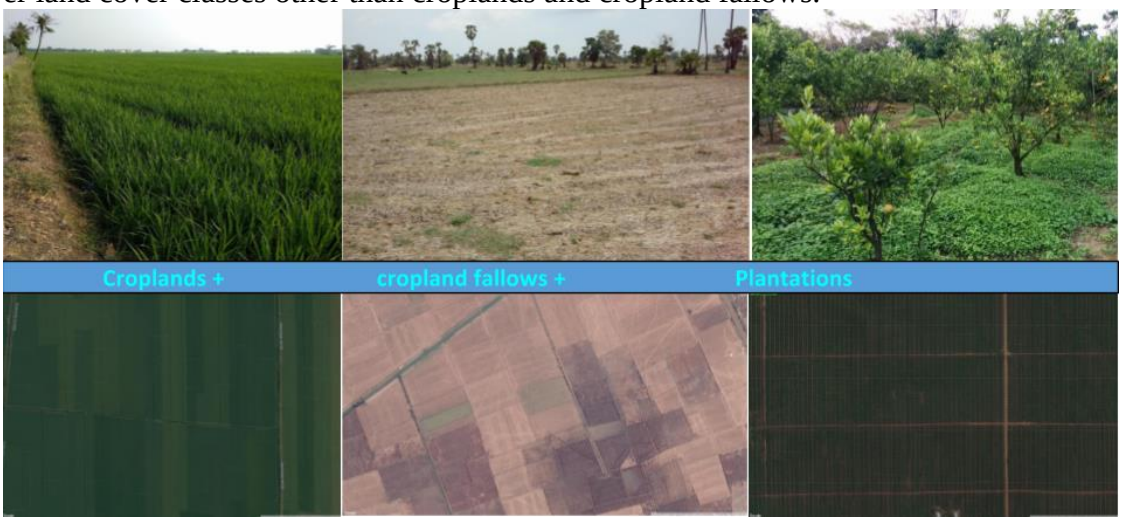

Figure 4. Illustration of definition of cropland mapping. Croplands included: (a) standing crop, (b) cropland fallows, and (c) permanent plantation crops. Note: + sign means adding. Means total net croplands= standing crops + cropland fallows + plantations.

ii. Algorithm

The study used one machine-learning algorithm to create the cropland extent product, which is the pixel-based supervised classifier Random Forest (RF). The algorithm is described in detail below. Afghanistan was stratified into two separate refined FAO agro-ecological zones and South Asia stratified into five zones (Figure 2) to facilitate the optimal classification.

c. Practical Description

i. Random Forest Classifier (RF)

The Random Forest classifier is more robust, relatively faster, and easier to implement than many other classifiers (Pelletier et al., 2016). The Random Forest classifier uses bootstrap aggregating (bagging) to form an ensemble of decision trees by searching random subspaces from

the given data (features) and the best splitting of the nodes by minimizing the correlation between the trees.

All supervised pixel-based classifications rely heavily on the input training samples. To discriminate croplands under various environments and condition, the sample size of the initial training dataset needs to be large, especially in complex regions. All samples were selected to represent a 90-m x 90-m polygon. First, we made extensive field campaigns in South Asia during the 2014-2016 crop growing seasons when data were collected on precise cropland locations as well as non-cropland locations. This effort led to collection of more than 2800 samples spread across South Asia. Second, we absorbed the ground data from previous efforts for South Asia region and other reliable sources. Third, sub-meter to 5-m very high spatial resolution imagery, available for us for the entire study region, was used to generate croplands *versus* non-cropland interpretations by multiple analyses across South Asia, Afghanistan and Iran and a total of ~5100 data samples were used from these interpretations. To move forward with a larger sample size, an iterative sample selection procedure was introduced with the following steps for training the Random Forest (RF) machine-learning algorithm as illustrated in Figure 1.

1. Build Random Forest classifier using existing training samples. Initially we start with a small number of samples and slowly increase the sample size till we reach high degree of accuracy and the accuracy plateaus at certain sample size;
2. Based on established classifier, classify 30-m MFDC using Random Forest algorithm in GEE cloud;
3. Visual assessment of classification results are compared with existing reference maps as well as sub-meter to 5-m very high spatial resolution imagery (VHRI); The process (Figure 1) was iterated until sufficient correspondence is achieved;
4. Added (see Figure 1) 'crop' samples in missing area and 'non-crop' samples by referencing sub-meter to 5-m very high spatial imagery from Google Earth Imagery. For cases hard to tell by interpretation (fallow-land or abandoned fields), historical Landsat Images and MODIS NDVI time-series are also referenced. All the samples selected to represent a 90-m x 90-m polygon.
5. Loop step 1-4 with enlarged training dataset until classification becomes stable.

The number of iterations required for the training sample selection is a function of the complexity of the area. The whole study area was divided into six zones; to carry out classification (Figure 2): the iterative selection will have to loop ~4-5 times to improve the initial classified results.

ii. Programming and codes

The pixel-based supervised machine-learning algorithm (RF) was coded on GEE using Python and Java Scripts using Application Programming Interface (API). The codes are made available in a zip file and are available for download along with this ATBD.

iii. Results

The machine learning algorithms (RF), discussed in previous sections, were trained to separate croplands *versus* non-croplands for each of the zones (Figure 2) based on knowledge generated using reference data. The machine learning algorithms were then run on the GEE cloud-computing environment using a Landsat-8 collection for each of the zones to separate croplands *versus* non-croplands. The process was iterated and knowledge in the algorithms

tweaked several times, before getting accurate results of croplands *versus* non-croplands. This process led to producing the Global Food Security-support Analysis Data @ 30-m cropland extent product for South Asia, Afghanistan and Iran (Figure 5). This product is available through the Land Processes Distributed Active Archive Center (LP DAAC). The same dataset is also available for visualization at https://croplands.org/app/map.

Zoom-in views show complete resolution of the imagery that shows individual farms (Figure 5). Full resolution of 30-m cropland extent can be visualized in croplands.org by zooming-in to specific areas as illustrated in lower panel (a) and (b) of Figures 5, and 6. For any area in South Asia or Afghanistan or Iran croplands can be visualized by zooming into specific areas in croplands.org. The background sub-meter to 5-m imagery, available for the regions on Google Earth helps evaluate the precision of the cropland extent product ("zoom in" and "toggle" cropland "on" and "off" to see the sub-meter to 5-m imagery in the background).

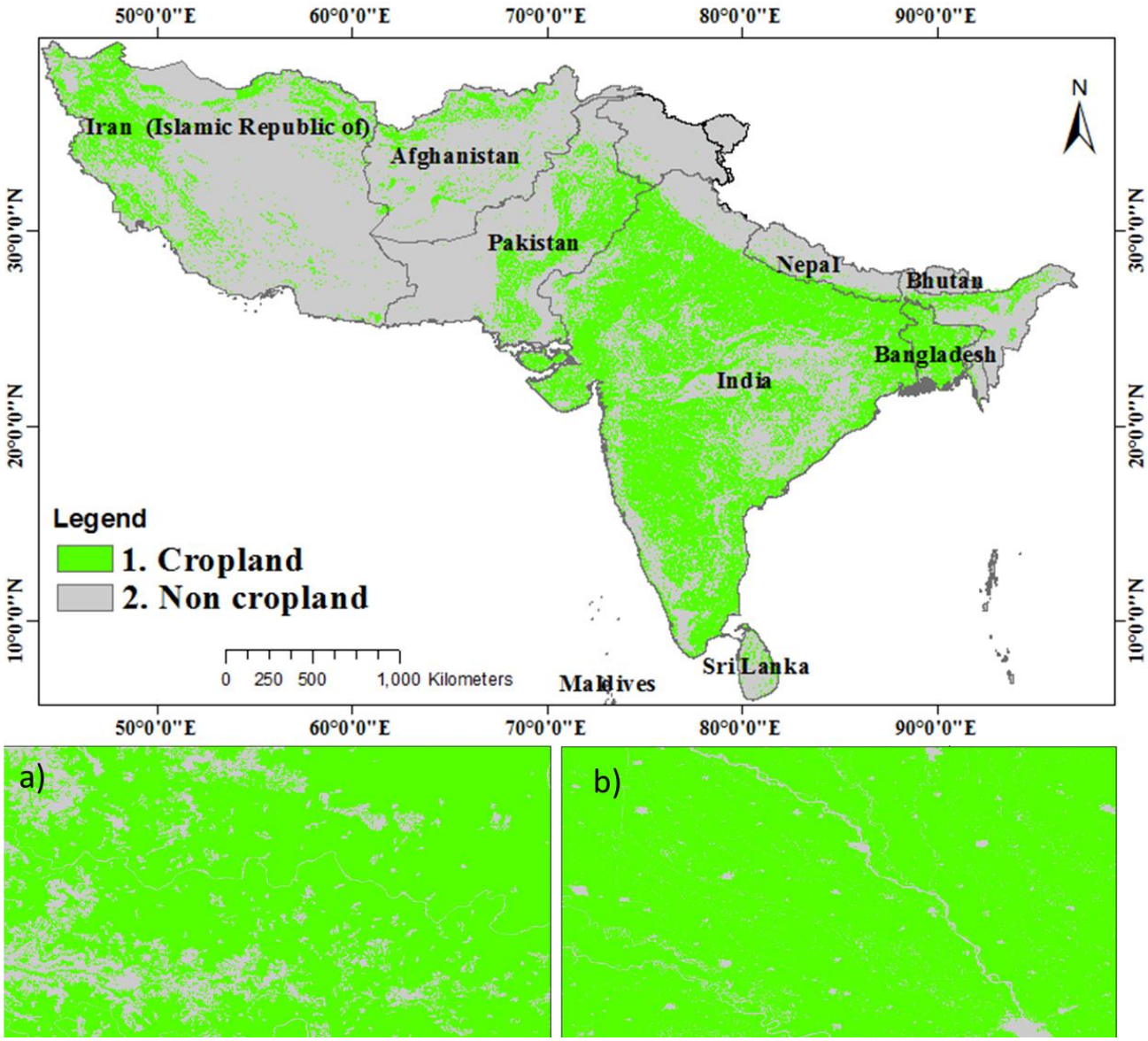

Figure 5. Cropland Extent Product at 30-m for South Asia (left image) with illustrative zoom in view for a location (bottom). This product is made available for visualization @: croplands.org. The data are downloadable from LP DAAC.

iv. Cropland Areas

Countrywide cropland areas calculated based on 30-m crop extent maps from this study are summarized here. Table 5 shows country-by-country cropland area statistics of all countries generated from this study and compared with several other sources such as the national census data based MIRCA2000 (Siebert *and* Portmann, 2010) which were also updated in the year 2015, The Food and Agricultural Organization (FAO) of United Nation's compiled statistics, MODIS 500-m derived cropland areas from GRIPC (Salmon et al., 2015), and GIAM-GMRCA (Thenkabail et al., 2009 and Biradar et al., 2009) derived cropland areas. Overall as per GFSAD30SAAFGIRCE estimates, total net cropland area of India is 180 Mha, which is top ranked country in the world in terms of cropland area. Net cropland area of Iran is 33 Mha followed by Pakistan with 21 Mha, Afghanistan 8.5 Mha, Nepal 2.0 Mha, Sri Lanka 1.5Mha, Bhutan 0.05Mha. Total 255 Mha from 9 countries of this study area.

Table 4. Net cropland areas (NCAs) derived based on 30-m GFSAD30 cropland product and comparison with other cropland products.

Country	**Land Area[1]**	**GFSAD30[2]**	**MIRCA 2014[3]**	**FAO Agri-cultural land[4]**	**GIAM-GMRCA[5]**	**GRIPC 2005[6]**
Name	**Ha**	**Ha**	**Ha**	**Ha**	**Ha**	**Ha**
India	297459504	179800110	177397578	169705109	150059162	187497499
Iran	162802013	33063882	16644983	18969365	8133031	7358862
Pakistan	77067449	21098899	25160408	21286800	17678708	20402689
Bangladesh	13014225	9562059	10027180	8545166	7771342	9234489
Afghanistan	65249570	8499171	9480926	7923190	3756220	909683
Nepal	14358108	2073750	3421684	2520250	4383047	3546940
Sri Lanka	6274038	1477022	2155484	2168910	2387275	3536804
Bhutan	3840909	46252	166573	99879	155065	99645
Maldives	29963	630		7000		

Note:

1= Total land area is land area excluding area under inland water bodies

2= GFSAD30 current study

3= Monthly irrigated and rainfed crop areas (MIRCA) around the year 2014 derived by Portman et al.

4= FAO Agricultural land area excluding pasture based on FAO2013 statistics consider nominal 2015

http://www.fao.org/faostat/en/#data/QC

5= Global croplands derived from Global Irrigated Area Mapping (GIAM) and Global Map of Rainfed Cropland Areas (GMRCA) by Thenkabail et al., 2009 and Biradar et al., 2009

6= Global rain-fed, irrigated, and paddy croplands (GRIPC) derived by solmon et al., 2015

V. Calibration Needs/Validation Activities

An independent accuracy assessment was performed each of six zones in the study area. For this assessment, 1500 validation samples were used to determine the accuracy of the final cropland extent map of South Asia, Afghanistan and Iran. An error matrix (Table 5) was generated for each of six zones providing producer's, user's, and overall accuracies (Story and Congalton, 1986, Congalton, 1991, and Congalton and Green, 2009).

For the entire study region all of 6 zones combined; the overall accuracies were 84.5% with producer's accuracy of 74.8% and user's accuracy of 82.0% (Table 6) for cropland class.

When considering all 6 zones, the overall accuracies ranged between 76% -96%, producer's accuracies ranged between 71-85% except for zone 2 (33.3%) and zone 6(48.4%); and user's accuracies ranged between 56-86% (Table 5). Zones that included a larger proportion of croplands had high overall, user's, and producer's accuracies. These results clearly imply the high level of confidence in differentiating croplands from non-croplands for the South Asia region.

Table 5. Independent Accuracy Assessment error matrix of 30-m Cropland Extent Product by Zone for the entire study Area (six zones).

Zone#1		Reference Data: Crop	Reference Data: No-Crop	Total	User Accuracy
Map Data	Crop	42	32	74	56.76%
	No-Crop	17	159	176	90.34%
Total		59	191	250	
Producer Accuracy		71.19%	83.25%		**80.40%**

Zone#2		Reference Data: Crop	Reference Data: No-Crop	Total	User Accuracy
Map Data	Crop	4	3	7	57.14%
	No-Crop	8	235	243	96.71%
Total		12	238	250	
Producer Accuracy		33.33%	98.74%		**95.60%**

Zone#3		Reference Data: Crop	Reference Data: No-Crop	Total	User Accuracy
Map Data	Crop	111	19	130	85.38%
	No-Crop	40	80	120	66.67%
Total		151	99	250	
Producer Accuracy		73.51%	80.81%		**76.40%**

Zone#4		Reference Data: Crop	Reference Data: No-Crop	Total	User Accuracy
Map Data	Crop	140	18	158	88.61%
	No-Crop	24	67	91	73.63%
Total		164	85	249	
Producer Accuracy		85.37%	78.82%		**83.13%**

Zone#5		Reference Data: Crop	Reference Data: No-Crop	Total	User Accuracy
Map Data	Crop	104	17	121	85.95%
	No-Crop	36	92	128	71.88%
Total		140	109	249	
Producer Accuracy		74.29%	84.40%		**78.71%**

Zone#6		Reference Data: Crop	Reference Data: No-Crop	Total	User Accuracy
Map Data	Crop	15	3	18	83.33%
	No-Crop	16	216	232	93.10%
Total		31	219	250	
Producer Accuracy		48.39%	98.63%		**92.40%**

Combined		Reference Data: Crop	Reference Data: No-Crop	Total	User Accuracy
Map Data	Crop	418	92	510	81.96%
	No-Crop	141	849	990	85.76%
Total		559	941	1,500	
Producer Accuracy		74.78%	90.22%		**84.47%**

VI. Constraints and Limitations

GFSAD30SAAFGIRCE product mapped the croplands of South Asia, Afghanistan and Iran @ nominal 30-m, which is the best known resolution for cropland mapping over such a large Agriculture area covering 9 countries. It also has high levels of accuracies with overall accuracies of 84.5%, Producer's accuracy of 74.8% and user's accuracy of 82.0% for cropland class.

A producer's accuracy of 74.8% for the cropland class means an error of omission of 25.2%, which suggests that roughly a quarter of the croplands were missing in the product. A user accuracy of 82.0 % for the cropland class means there is an error of commission of 18.0 %, meaning less than 20.0% of non-croplands are mapped as croplands. We tweaked the machine learning algorithms (section IV) to maximize capturing as much croplands as feasible automatically. In this process, some non-croplands are mapped as croplands as well. This is a preferred solution, in order to not miss croplands or to only miss them minimally. As a compromise, mapping some non-croplands as croplands becomes unavoidable.

Numerous issues cause uncertainties and limitations in cropland extent product. Some of these issues are discussed here. First, temporal coverage. The 16 day Landsat-8 and 16-day Landsat-7 coverage when put together, lead to substantial temporal coverage. Yet, if we look at Figure 3, we were only able to achieve seasonal (and not bi-weekly or monthly) cloud-free or near cloud-free mosaics of the entire study area. This is not surprising given such a large area involved and frequent cloud across the study area. As a result, if we were to have daily coverage over an area (e.g., like MODIS) then it becomes feasible to have more frequent (e.g., monthly or bi-monthly composites) temporal coverage of the continent that will help advance cropland mapping at improved accuracies. Currently, even with Landsat 8 satellite, at best, we have two images per month (compared to 30 images of MODIS when we consider daily daytime coverage of MODIS). Second, is the limitation of the reference training and validation data. In this project, we already have large training and validation data compared to any previous work as described in various previous sections. Nevertheless, much wider and extensive field visits to different parts of the study regions will be helpful in better understanding of the issues involved and as a result better mapping. We had extensive field visits in India, Bangladesh, Sri Lanka, but these data were mostly acquired one time. The greatest difficulties in cropland mapping in eastern part of India were in mountainous agriculture (e.g., terrace agriculture), cropland fallows (e.g., whether a fallow is 1 year or 5-year or permanent). These and numerous other issues (e.g., implementing machine learning algorithms and uncertainties inherent in them) will continue to be there in cropland mapping over such large agriculture areas in china. Nevertheless advances made in this study is significant, especially in developing a nominal 30-m cropland extent of a large agriculture countries like India, Iran Pakistan and Sri Lanka at very good accuracies.

VII. Publications

Congalton, R.G., Gu, J., Yadav, K., Thenkabail, P.S., and Ozdogan, M. 2014. Global Land Cover Mapping: A Review and Uncertainty Analysis. Remote Sensing Open Access Journal. Remote Sens. 2014, 6, 12070-12093; http://dx.doi.org/10.3390/rs61212070.

Congalton, R.G, 2015. Assessing Positional and Thematic Accuracies of Maps Generated from Remotely Sensed Data. Chapter 29, In Thenkabail, P.S., (Editor-in-Chief), 2015. "Remote Sensing Handbook" Volume I: Volume I: Data Characterization, Classification, and Accuracies: Advances of Last 50 Years and a Vision for the Future. Taylor and Francis Inc.\CRC Press, Boca Raton, London, New York. Pp. 900+. In Thenkabail, P.S., (Editor-in-Chief), 2015. "Remote Sensing Handbook" Volume I: **): Remotely Sensed Data Characterization, Classification, and Accuracies**. Taylor and Francis Inc.\CRC Press, Boca Raton, London, New York. ISBN 9781482217865 - CAT# K22125. Print ISBN: 978-1-4822-1786-5; eBook ISBN: 978-1-4822-1787-2. Pp. 678.

Gumma, M.K., Thenkabail, P.S.,Teluguntla, P., Rao, M.N., Mohammed, I.A., and Whitbread, A.M. 2016. Mapping rice-fallow cropland areas for short-season grain legumes intensification in South Asia using MODIS 250 m time-series data. International Journal of Digital Earth, http://dx.doi.org/10.1080/17538947.2016.1168489

Massey, R., Sankey, T.T., Congalton, R.G., Yadav, K., Thenkabail, P.S., Ozdogan, M., Sánchez Meador, A.J. 2017. MODIS phenology-derived, multi-year distribution of conterminous U.S. crop types, Remote Sensing of Environment, Volume 198, 1 September 2017, Pages 490-503, ISSN 0034-4257, https://doi.org/10.1016/j.rse.2017.06.033.

Phalke, A. R., Ozdogan, M., Thenkabail, P. S., Congalton, R. G., Yadav, K., & Massey, R. et al. (2017). A Nominal 30-m Cropland Extent and Areas of Europe, Middle-east, Russia and Central Asia for the Year 2015 by Landsat Data using Random Forest Algorithms on Google Earth Engine Cloud. (in preparation).

Portmann, F.T., Siebert, S., Döll, P., 2010. MIRCA2000—Global monthly irrigated and rainfed crop areas around the year 2000: A new high-resolution data set for agricultural and hydrological modeling. Global Biogeochemical Cycles 24(1).

Teluguntla, P., Thenkabail, P.S., Xiong, J., Gumma, M.K., Congalton, R.G., Oliphant, A., Poehnelt, J., Yadav, K., Rao, M., and Massey, R. 2017. Spectral matching techniques (SMTs) and automated cropland classification algorithms (ACCAs) for mapping croplands of Australia using MODIS 250-m time-series (2000–2015) data, International Journal of Digital Earth. DOI:10.1080/17538947.2016.1267269.IP-074181, http://dx.doi.org/10.1080/17538947.2016.1267269.

Teluguntla, P., Thenkabail, P., Xiong, J., Gumma, M.K., Giri, C., Milesi, C., Ozdogan, M., Congalton, R., Yadav, K., 2015. CHAPTER 6 - Global Food Security Support Analysis Data at Nominal 1 km (GFSAD1km) Derived from Remote Sensing in Support of Food Security in the Twenty-First Century: Current Achievements and Future Possibilities, in: Thenkabail, P.S. (Ed.), Remote Sensing Handbook (Volume II): Land Resources Monitoring, Modeling, and

Mapping with Remote Sensing. CRC Press, Boca Raton, London, New York., pp. 131–160. Link.

Xiong, J., Thenkabail, P.S., Tilton, J.C., Gumma, M.K., Teluguntla, P., Oliphant, A., Congalton, R.G., Yadav, K. 2017. A Nominal 30-m Cropland Extent and Areas of Continental Africa for the Year 2015 by Integrating Sentinel-2 and Landsat-8 Data using Random Forest, Support Vector Machines and Hierarchical Segmentation Algorithms on Google Earth Engine Cloud. Remote Sensing Open Access Journal (in review).

Xiong, J., Thenkabail, P.S., Gumma, M.K., Teluguntla, P., Poehnelt, J., Congalton, R.G., Yadav, K., Thau, D. 2017. Automated cropland mapping of continental Africa using Google Earth Engine cloud computing, ISPRS Journal of Photogrammetry and Remote Sensing, Volume 126, April 2017, Pages 225-244, ISSN 0924-2716, https://doi.org/10.1016/j.isprsjprs.2017.01.019.

Web sites and Data portals:
http://croplands.org (30-m global croplands visualization tool)
http://geography.wr.usgs.gov/science/croplands/index.html (GFSAD30 web portal and dissemination)
http://geography.wr.usgs.gov/science/croplands/products.html#LPDAAC (dissemination on LP DAAC)
http://geography.wr.usgs.gov/science/croplands/products.html (global croplands on Google Earth Engine)
croplands.org (crowdsourcing global croplands data)

VIII. Acknowledgements

The project was funded by the National Aeronautics and Space Administration (NASA) grant number: NNH13AV82I through its MEaSUREs (Making Earth System Data Records for Use in Research Environments) initiative. The United States Geological Survey (USGS) provided supplemental funding from other direct and indirect means through its Land Change Science (LCS), and Land Remote Sensing (LRS) programs as well as its Climate and Land Use Change Mission Area. The project was led by United States Geological Survey (USGS) in collaboration with NASA AMES, the University of New Hampshire (UNH), the California State University Monterey Bay (CSUMB), the University of Wisconsin (UW), NASA GSFC, and the Northern Arizona University. There were a number of International partners including The International Crops Research Institute for the Semi-Arid Tropics (ICRISAT).

IX. Contact Information

LP DAAC User Services
U.S. Geological Survey (USGS)
Center for Earth Resources Observation and Science (EROS)
47914 252nd Street
Sioux Falls, SD 57198-0001

Phone Number: 605-594-6116
Toll Free: 866-573-3222 (866-LPE-DAAC)
Fax: 605-594-6963
Email: lpdaac@usgs.gov
Web: https://lpdaac.usgs.gov

For the Principal Investigators, feel free to write to:
Prasad S. Thenkabail at pthenkabail@usgs.gov or
Pardhasaradhi Teluguntla at pteluguntla@usgs.gov or
Adam Oliphant at aoliphant@usgs.gov
More details about the GFSAD30 project and products can be found at:
https://globalcroplands.org

X. Citations

M.K.Gumma, P. Thenkabail, P. Teluguntla, A. Oliphant, J. Xiong, R. Congalton, K. Yadav, C. Smith (2017). NASA Making Earth System Data Records for Use in Research Environments (MEaSUREs) Global Food Security-support Analysis Data (GFSAD) 30-m Cropland Extent 2015 South Asia, Afghanistan and Iran 30m (GFSAD30SAAFGIRCE). NASA EOSDIS Land Processes DAAC. Retrieved from
https://doi.org/10.5067/MEaSUREs/GFSAD/GFSAD30SAAFGIRCE.001

XI. References

Biradar, C.M., Thenkabail, P.S., Noojipady, P., Li, Y., Dheeravath, V., Turral, H., Velpuri, M., Gumma, M.K., Gangalakunta, O.R.P., & Cai, X.L. 2009. A global map of rainfed cropland areas GMRCA at the end of last millennium using remote sensing. International Journal of Applied Earth Observation and Geoinformation, 11, 114-129

Büttner, G., 2014. CORINE Land Cover and Land Cover Change Products, in: Land Use and Land Cover Mapping in Europe. Springer Netherlands, Dordrecht, pp. 55–74.

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

Congalton, R. and K. Green. 2009. Assessing the Accuracy of Remotely Sensed Data: Principles and Practices. 2nd Edition. CRC/Taylor & Francis, Boca Raton, FL 183p

Congalton, R.G. 2015. Assessing positional and thematic accuracies of maps generated from remotely sensed data."Remote Sensing Handbook" three-volume set: Remotely Sensed Data Characterization, classification, and accuracies, Taylor and Francis Inc.\CRC Press, Boca Raton, London, New York. Pp. 800+. Pp. 625-662.
http://www.fao.org/faostat/en/#data/QC

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D. and Moore, R., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sensing of Environment.
Huang, C., Davis, L.S., Townshend, J.R.G., 2010. An assessment of support vector machines for land cover classification. International Journal of Remote Sensing 23, 725–749.
Irons, J.R., Dwyer, J.L., Barsi, J.A., 2012. The next Landsat satellite: The Landsat Data Continuity Mission. Remote Sensing of Environment 122, 11–21.

Latifovic, R., Homer, C., Ressl, R., Pouliot, D., Hossain, S.N., Colditz, R.R., Olthof, I., Giri, C. and Victoria, A., 2010. North American Land Change Monitoring System (NALCMS). Remote sensing of land use and land cover: principles and applications. CRC Press, Boca Raton.

Lillesand, T., Kiefer, R.W. and Chipman, J., 2014. Remote sensing and image interpretation. John Wiley & Sons.

Myint, S.W., Myint, S.W., Gober, P., Brazel, A., Gober, P., Brazel, A., Grossman-Clarke, S., Grossman-Clarke, S., Weng, Q., 2011. Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery. Remote Sensing of Environment 115, 1145–1161.

Pelletier, C., Valero, S., Inglada, J., Champion, N., Dedieu, G., 2016. Assessing the robustness of Random Forests to map land cover with high resolution satellite image time series over large areas. Remote Sensing of Environment 187, 156–168.

Portmann, F.T., Siebert, S., & Döll, P. 2010. MIRCA2000—Global monthly irrigated and rain-fed crop areas around the year 2000: A new high-resolution data set for agricultural and hydrological modeling. Global biogeochemical cycles, 24

Roy, D.P., Wulder, M.A., Loveland, T.R., C E, W., Allen, R.G., Anderson, M.C., Helder, D., Irons, J.R., Johnson, D.M., Kennedy, R., Scambos, T.A., Schaaf, C.B., Schott, J.R., Sheng, Y., Vermote, E.F., Belward, A.S., Bindschadler, R., Cohen, W.B., Gao, F., Hipple, J.D., Hostert, P., Huntington, J., Justice, C.O., Kilic, A., Kovalskyy, V., Lee, Z.P., Lymburner, L., Masek, J.G., McCorkel, J., Shuai, Y., Trezza, R., Vogelmann, J., Wynne, R.H., Zhu, Z., 2014. Landsat-8: Science and product vision for terrestrial global change research. Remote Sensing of Environment 145, 154–172.

Salmon, J.M., Friedl, M.A., Frolking, S., Wisser, D., & Douglas, E.M. 2015. Global rain-fed, irrigated, and paddy croplands: A new high resolution map derived from remote sensing, crop inventories and climate data. International Journal of Applied Earth Observation and Geoinformation, 38, 321-334

Shi, D., Yang, X., 2015. Support Vector Machines for Land Cover Mapping from Remote Sensor Imagery, in: Monitoring and Modeling of Global Changes: A Geomatics Perspective. Springer Netherlands, Dordrecht, pp. 265–279.

Story, M. and R. Congalton. 1986. Accuracy assessment: A user's perspective. Photogrammetric Engineering and Remote Sensing. Vol. 52, No. 3. pp. 397-399.

Teluguntla, P., Thenkabail, P., Xiong, J., Gumma, M.K., Giri, C., Milesi, C., Ozdogan, M., Congalton, R., Yadav, K., 2015. CHAPTER 6 - Global Food Security Support Analysis Data at Nominal 1 km (GFSAD1km) Derived from Remote Sensing in Support of Food Security in the Twenty-First Century: Current Achievements and Future Possibilities, in: Thenkabail, P.S. (Ed.), Remote Sensing Handbook (Volume II): Land Resources Monitoring, Modeling, and Mapping with Remote Sensing. CRC Press, Boca Raton, London, New York., pp. 131–160.

Thenkabail, P.S., Biradar, C.M., Noojipady, P., Dheeravath, V., Li, Y., Velpuri, M., Gumma, M., Gangalakunta, O.R.P., Turral, H., & Cai, X. 2009b. "Global irrigated area map GIAM, derived from remote sensing, for the end of the last millennium." International Journal of Remote Sensing, 30, 3679-3733

Thenkabail, P.S., Hanjra, M. a, Dheeravath, V., Gumma, M., 2010. A Holistic View of Global Croplands and Their Water Use for Ensuring Global Food Security in the 21st Century through Advanced Remote Sensing and Non-remote Sensing Approaches. Remote Sensing 2, 211.
Tian, S., Zhang, X., Tian, J., Sun, Q., 2016. Random Forest Classification of Wetland Landcovers from Multi-Sensor Data in the Arid Region of Xinjiang, China. Remote Sensing 8, 954.