

Agricultural cropland extent and areas of South Asia derived using Landsat satellite 30-m time-series big-data using random forest machine learning algorithms on the Google Earth Engine cloud

Murali Krishna Gumma, Prasad S. Thenkabail, Pardhasaradhi G. Teluguntla, Adam Oliphant, Jun Xiong, Chandra Giri, Vineetha Pyla, Sreenath Dixit & Anthony M Whitbread

To cite this article: Murali Krishna Gumma, Prasad S. Thenkabail, Pardhasaradhi G. Teluguntla, Adam Oliphant, Jun Xiong, Chandra Giri, Vineetha Pyla, Sreenath Dixit & Anthony M Whitbread (2019): Agricultural cropland extent and areas of South Asia derived using Landsat satellite 30-m time-series big-data using random forest machine learning algorithms on the Google Earth Engine cloud, GIScience & Remote Sensing, DOI: [10.1080/15481603.2019.1690780](https://doi.org/10.1080/15481603.2019.1690780)

To link to this article: <https://doi.org/10.1080/15481603.2019.1690780>



© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 22 Nov 2019.



[Submit your article to this journal](#)



Article views: 983



[View related articles](#)



[View Crossmark data](#)

Agricultural cropland extent and areas of South Asia derived using Landsat satellite 30-m time-series big-data using random forest machine learning algorithms on the Google Earth Engine cloud

Murali Krishna Gumma ^a, Prasad S. Thenkabail ^b, Pardhasaradhi G. Teluguntla ^{b,c}, Adam Oliphant ^b, Jun Xiong ^b, Chandra Giri ^b, Vineetha Pyla ^d, Sreenath Dixit ^a and Anthony M Whitbread ^a

^aRS/GIS Lab, Innovation Systems for the Drylands, International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru 502 324, India; ^bWestern Geographic Science Center, U.S. Geological Survey (USGS), Flagstaff, AZ, USA; ^cBay Area Environmental Research Institute (BAERI), NASA Research Park, Moffett Field, CA, USA; ^dJawaharlal Nehru Technological University (JNTU), Hyderabad, India

ABSTRACT

The South Asia (India, Pakistan, Bangladesh, Nepal, Sri Lanka and Bhutan) has a staggering 900 million people (~43% of the population) who face food insecurity or severe food insecurity as per United Nations, Food and Agriculture Organization's (FAO) the Food Insecurity Experience Scale (FIES). The existing coarse-resolution (≥ 250 -m) cropland maps lack precision in geo-location of individual farms and have low map accuracies. This also results in uncertainties in cropland areas calculated from such products. Thereby, the overarching goal of this study was to develop a high spatial resolution (30-m or better) baseline cropland extent product of South Asia for the year 2015 using Landsat satellite time-series big-data and machine learning algorithms (MLAs) on the Google Earth Engine (GEE) cloud computing platform. To eliminate the impact of clouds, 10 time-composited Landsat bands (blue, green, red, NIR, SWIR1, SWIR2, Thermal, EVI, NDVI, NDWI) were derived for each of the three time-periods over 12 months (monsoon: Days of the Year (DOY) 151–300; winter: DOY 301–365 plus 1–60; and summer: DOY 61–150), taking the every 8-day data from Landsat-8 and 7 for the years 2013–2015, for a total of 30-bands plus global digital elevation model (GDEM) derived slope band. This 31-band mega-file big data-cube was composed for each of the five agro-ecological zones (AEZ's) of South Asia and formed a baseline data for image classification and analysis. Knowledge-base for the Random Forest (RF) MLAs were developed using spatially well spread-out reference training data (N = 2179) in five AEZs. The classification was performed on GEE for each of the five AEZs using well-established knowledge-base and RF MLAs on the cloud. Map accuracies were measured using independent validation data (N = 1185). The survey showed that the South Asia cropland product had a producer's accuracy of 89.9% (errors of omissions of 10.1%), user's accuracy of 95.3% (errors of commission of 4.7%) and an overall accuracy of 88.7%. The National and sub-national (districts) areas computed from this cropland extent product explained 80–96% variability when compared with the National statistics of the South Asian Countries. The full-resolution imagery can be viewed at full-resolution, by zooming-in to any location in South Asia or the world, at www.croplands.org and the cropland products of South Asia downloaded from The Land Processes Distributed Active Archive Center (LP DAAC) of National Aeronautics and Space Administration (NASA) and the United States Geological Survey (USGS): <https://lpdaac.usgs.gov/products/gfsad30saafgircsev001/>.

ARTICLE HISTORY

Received 2 May 2019
Accepted 4 November 2019

KEYWORDS

Google Earth Engine;
random forest; Landsat;
cloud computing; 30m-South
Asia croplands

1. Introduction

Up-to-date spatial distribution of cropland maps that are precise, accurate, and at high-spatial resolution (30-m or better) is of great importance for sustainable agriculture production and food security (Teluguntla et al. 2015; Thenkabail et al. 2012). However, cropland areas are changing rapidly over time across South Asia due to climate variability, rural migration to urban areas, industrialization, population growth, and other socio-economic issues. Modern sustainable agriculture and sustained rural development are critical to eradication of poverty and improving livelihoods. Of the nearly

1.9 billion population in South Asia, about 250 million people (~15% of the population) still live under the international poverty line of \$1.90 purchasing power parity (PPP) and a staggering about 900 million (~48% of the population) in lower middle income class value of US\$3.20 PPP (WorldBank 2018)

Agriculture is still the main livelihood of most South Asian population and much of this is small-holder dominant. For example, in India, 80% of farmers have less than 2 Ha and in Pakistan, 58% of farmers have less than 2 Ha. In Bangladesh and Nepal, over 90% of farmers have less than 1 Ha. In Bhutan, the average farm size is slightly

over 1 Ha and women own about 70% of total agricultural land. South Asia has one of the largest cropland areas of the world, both in terms of the net cropland area as well as the gross cropland area (Teluguntla et al. 2015; Thenkabail et al. 2012). Both irrigation and rainfed agriculture are prominent in South Asia and comprise 35% and 65% of the total cropland area, respectively; South Asia also has one of the highest irrigated areas of the world (Thenkabail et al. 2009b).

The population of the region is projected to grow faster than its ability to produce sufficient food to meet demand, especially rice and wheat which are the two major staple crops (Papademetriou 2000). Subsequently, this could make South Asia food insecure in the coming decades. To evaluate and monitor this situation, high-resolution (30-m or better) cropland maps with accurate spatial distribution are required for all higher-level products (e.g. crop type mapping, crop productivity, and crop water productivity modeling and mapping, biomass and yield assessments). This will provide critical information to planners, decision-makers, and scientists on where exactly crops are cultivated and the dynamics affecting them over space and time. High-resolution cropland maps are capable of analyzing and interpreting precise and accurate assessments of agricultural water use and irrigation systems. This is extremely important because about 80% of all human water use currently goes toward producing food. Furthermore, cropland extent maps at high-resolution help establish cropland area statistics at individual farm level to various administrative units. Overall, the importance of a detailed high-resolution cropland extent maps in food and water security studies cannot be over-emphasized (Thenkabail et al. 2010; Teluguntla et al. 2015; Waldner et al. 2015, 2016).

In South Asia there has been large disagreement between available cropland map products (Fritz and See 2008; Giri, Zhu, and Reed 2005; Herold et al. 2008; Roy et al. 2015) with cropland area as well as geo-precision of cropland location (Dheeravath et al. 2010; Gumma et al. 2011a, 2016; Teluguntla et al. 2015; Thenkabail and Wu 2012). These differences are attributed to factors such as fragmented and heterogeneous rural landscapes (Biggs et al. 2006; Gray et al. 2014), coarse spatial resolution of the imagery (Thenkabail et al. 2009a), class definition issues (Teluguntla et al. 2015), methodological issues, such as many decisions on the most appropriate methods and thresholds used in those methods (Gumma et al. 2014). Additional challenges of mapping croplands in South Asia also include: (a) small farm sizes, (b) fragmented farms (scattered agriculture), (c) spectral similarity with grasslands and fallows in arid and semi-arid areas (Dheeravath et al. 2010; Thenkabail, Schull, and Turrall 2005), (d) high regional variability in terms of agricultural systems and calendars between the desert margins and other agro-ecological zones (Roy et al. 2015),

(e) mixed cropping with multiple crops in same farm (Giri, Zhu, and Reed 2005), and (f) mountainous terrain (i.e. Nepal and Sri Lanka).

Time series analysis of satellite images have been used widely for croplands monitoring (Biggs et al. 2006; Dheeravath et al. 2010; Dong et al. 2015; Gumma et al. 2011a, 2018; Thenkabail et al. 2007a, 2009b). However, most of these studies have been processed as standalone products using powerful, yet, time-consuming methods and approaches that were required to achieve the best possible results. Several advances in semi-automation were attempted by a number of studies mapping cropland extent over large areas (Pittman et al. 2010; Radoux et al. 2014; Salmon et al. 2015; Thenkabail and Wu 2012). These semi-automated methods also used data from a wide array of sensors, mapping croplands at various spatial resolutions of 250 to 10,000 m (Dheeravath et al. 2010; Gumma et al. 2011b, 2011c; Siebert et al. 2005; Thenkabail, Schull, and Turrall 2005; Vörösmarty et al. 2000). These include: semi-automated methods include phenology-based algorithms (Xiao et al. 2006); decision tree algorithms that were mainly based on monthly NDVI thresholds at different crop growth stages (Defries, Hansen, and Townshend 2000; Friedl and Brodley 1997; Otukey and Blaschke 2010; Pal and Mather 2003); spectral matching techniques (Teluguntla et al. 2017; Gumma et al. 2014; Thenkabail et al. 2007b) and other machine learning algorithms (Duro, Franklin, and Dubé 2012; Lary et al. 2016). More recently, significant advances have been made in mapping agricultural croplands using high-spatial resolution (30-m or better) time-series imagery over large areas such as continents using machine learning algorithms on Google Earth Engine (GEE) cloud-computing platforms (Oliphant et al. 2019; Teluguntla et al. 2018; Sangeeta, Gensuo, and Anzhi 2018; Xiong et al. 2017a, 2017b; Gorelick et al. 2017).

However, all of the above studies were constrained by the limited collection of in-field ground reference data so crucial for 1. development of accurate and precise knowledge-base for the machine learning algorithms, 2. Class identification and labeling, and 3. Class validation. This study overcomes this limitation of other studies by gathering very rich set of in-field ground reference training and validation data covering all the South Asian countries through several major field campaigns. Furthermore, this study had access to sub-national statistical data from all South Asian countries which helped validate sub-national area statistics derived from this study. Also, there is not a single high resolution (30-m or better) wall-to-wall seamless cropland product for entire South Asia; existing studies are often limited to small areas within South Asia or were conducted using coarser resolution data.

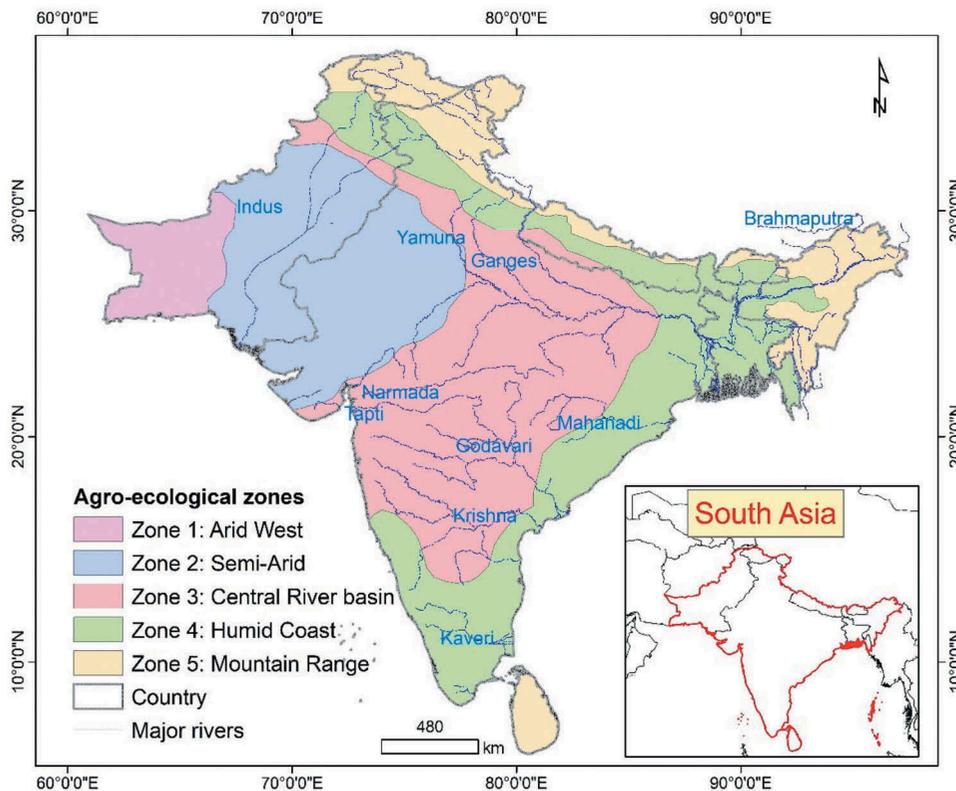


Figure 1. Study area and AEZs of South Asia. Study area of South Asia with five FAO simplified agro-ecological zones (AEZs).

The overarching goal of this research was to produce a precise and accurate cropland extent product of entire South Asia (Figure 1) using Landsat 30-m data with machine learning algorithms (MLAs) on the GEE cloud computing platform. Random forest (RF) MLA was used because it is an extremely effective classifier that is resistant to overfitting (Oliphant et al. 2019; Teluguntla et al. 2018; Xiong et al. 2017a; Rodriguez-Galiano et al. 2012). Extensive training and validation data spread across South Asia area were derived from various field data collections from many campaigns and were also sourced from sub-meter to 5 m very high-resolution imagery (VHRI). These reference data were used to train the RF MLAs in each of the five distinct agro-ecological zones (AEZs) of South Asia (Figure 1). The reference validation data, sourced from ground data as well as VHRI data, were used in developing accuracy error matrices for each of these five AEZs as well as for the entire South Asia. The goal of this study is also to help determine consistent and objective cropland areas at the national and sub-national level and compare these areas with the national and sub-national cropland records obtained from country statistics.

2. Study area

South Asia is located between 5°38' and 36°54' latitudes, 61°05' and 97°14' longitudes, with total geographical area of 477 Mha (Figure 1). This study includes 6 countries of

South Asia (Pakistan, India, Nepal, Bhutan, Bangladesh and Sri Lanka; Afghanistan and Maldives not included in the study). In South Asia, 80% of poor live in rural areas and are mostly depend on agriculture for their livelihood (WorldBank 2015). Nine major river basins were included in the study area: The Indus, Ganges, Brahmaputra, Narmada, Tapti, Godavari, Krishna, Kaveri and Mahanadi. South Asia is a major world leader in terms of the large scale of cropland areas as well as irrigated land. There are many major and minor irrigation projects in South Asia, covering a total command area of 133 Mha (Thenkabail et al. 2009a). The combination of irrigated and rainfed croplands and a wide variability in their crop types, cropping patterns, crop productivities, and crop dynamics make it complex to map croplands; requiring rich reference data to cut across these complexities over space and time. Rice and wheat are the two major crop in this region. Rice is grown two times a year in large parts of the region; in Bangladesh it is grown three times a year (Gumma et al. 2014).

Agricultural practices such as planting, harvesting, types of crops, and climate regimes vary greatly in South Asia, from generally dry arid climate in Western India and much of Pakistan to humid tropical climate in Southern India and Bangladesh. South Asia has five broad climate types: humid tropics, sub-humid tropics, semi-arid tropics, semi-arid, subtropics and arid (HarvestChoice 2009). To account for these differences in climate and

agricultural practices we divided South Asia into five Zones based on the FAO Agro-Ecological Zones or AEZ (Figure 1). Each AEZ zone was classified separately with separate training data and each zone was validated separately with independent validation data; these will be described in the methods section below.

3. Input data

3.1. Landsat 30-m time-series data

The input data used for cropland extent mapping over entire South Asia consisted of Landsat 8 and 7, 30-m time series data over 2 years (2014 and 2015) (Table 1). Landsat 30-m data was preferred because of their high-quality spectral calibrations and high spatial (30-m) and temporal (8 days from two satellites) resolution. The Landsat satellite series are launched by the NASA and the acquired data was freely processed and distributed by the USGS to the public globally.

After reviewing the literature and from earlier experimentation, we chose 10 bands for this analysis consisting of blue, green, red, near-infrared (NIR), SWIR1, SWIR2 and thermal, along with vegetation indices NDVI, EVI and LSWI for this classification (Table 1). NDVI (Normalized Difference Vegetation Index) was used for the identification of dense vegetation including forests. EVI (Enhanced Vegetation Index) is an optimized vegetation index which was used to detect the sensitive vegetation which are not clearly detected by NDVI. NDWI (Normalized Difference Water Index) was used for the identification of croplands. Images used were pre-processed to Top of the Atmosphere (TOA) by the USGS: Surface Reflectance images were not widely available when this study was conducted for Landsat 8 on GEE. Additionally, we decided to use the GDEM-derived elevation, helps providing classes which are otherwise same spectrally, into distinct classes in random forest classifier. This is specifically useful in classifying croplands in low elevation (e.g. rice in the river deltas) as opposed to croplands or non-croplands in uplands in higher elevations. Based on elevation, these spectrally

similar classes get classified into distinct classes which are then identified using ground data along with other data such as the sub-meter to 5-m data and ancillary data.

As Landsat produces optical data, there are significant cloud issues. In order to get cloud-free pixels over entire South Asia we composited 10 bands of Landsat (8 and 7 combined) data (Table 1) over 2 years (2014 and 2015) and three time periods (Table 2): Period 1, (Monsoon season: 151–300 DOY [day of year]); Period 2, (winter or rabi season: 301–365 DOY extending to 1–60 DOY); Period 3, (summer season: 61–150 DOY). For example, there will be 18 Landsat images over 150 days of period 1. So, for each band (e.g. blue band) a single time-composited layer is generated from the 18 layers of blue band. This ensures with almost certainty there will be a single pixel from 18 dates that is cloud free. The same process is applied for each of the 10 bands resulting in 10 time-composited bands that are almost completely cloud free over entire South Asia for the given period. A similar process is applied to obtain 10 time-composited bands for Period 2 and 3, respectively, with a combined total of 30 bands for the three Periods. These 30 Landsat-derived bands were merged with the slope GDEM band to create a 31-band mega-file data-cube (MGDC) (Table 2, Figure 2) for each of the five AEZs (Figure 1). By establishing these MFDC cloud free data stacks ensures that no erroneous classification results due to missing data occurred in portions of the image (Teluguntla et al. 2018). By overcompensating data over multiple years (2014 and 2015) it helps to overcome inter-annual variability. Since our goal was to obtain a cropland extent product for the nominal year 2015, we only use 2 years of data (2014 and 2015) that includes cloud-free pixels over nearly 100% of the South Asia, except for an insignificant percentile (<0.05%) of noisy pixels.

3.2. Reference training and validation: introduction

Reference training and validation data (Table 3) were gathered from two major sources:

Table 1. Characteristics of multi-temporal Landsat 8 and 7 data used in this study. Overall 10 bands of data (seven spectral bands plus three indices) were used for every 8 days (from two satellites) over 2014 and 2015.

Band name	Landsat 8 OLI Spectral range (µm)	Landsat 7 ETM+ Spectral range (µm)	Vegetation index (VI)	Equation
Blue	0.450–0.515	0.450–0.516	EVI	$EVI = 2.5 (NIR-red)/(NIR+6*red-7.5*blue+1)$
Green	0.525–0.600	0.525–0.601		
Red	0.630–0.680	0.630–0.681	NDWI	$NDWI = (NIR-SWIR1)/(NIR+SWIR1)$
NIR	0.845–0.885	0.845–0.886		
SWIR1	1.560–1.660	1.560–1.661		
SWIR2	2.100–2.300	2.100–2.301	NDVI	$NDVI = (NIR-red)/(NIR+red)$
Thermal	10.6–11.2	10.6–11.3		

NIR = near-infrared, SWIR = short-wave infrared, TIR = thermal infrared.

NDVI = normalized difference vegetation index, NDWI = normalized difference water index.

EVI = enhanced vegetation index.

Table 2. Composition of 31-band time-composited mega file data cubes (MFDCs) for each of the three periods (Period 1: Monsoon, Period 2: Winter, and Period 3: Summer) using 2 years (2014 and 2015) for each of the five AEZs.

Region name	Landsat image series	Years of data	Julian days over data are time-composited	Bands used ^a for each composited period	Total # of bands used in mega-file data cube	Data provider name
South Asia	Landsat-8 and 7	2014, 2015	Period 1 , Monsoon season: 151–300 days; Period 2 , winter or rabi season: 301–365, 1–60 Period 3 , summer season: 61–150 DOY	blue, green, red, NIR, SWIR-1, SWIR-2, TIR1, NDVI, EVI and NDWI (n = 10)	31 (10 bands 3 periods) + slope derived from GDEM)	USGS and NASA

aNIR = near-infrared, SWIR = short-wave infrared, TIR = thermal infrared.

NDVI = normalized difference vegetation index, NDWI = normalized difference water index.

EVI = enhanced vegetation index.

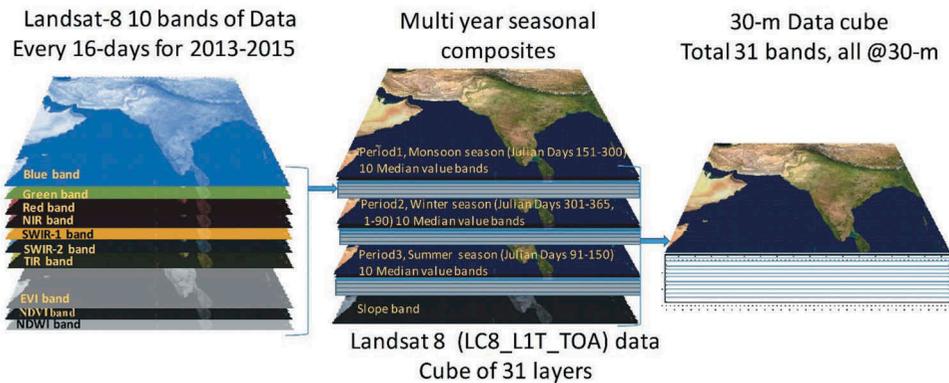


Figure 2. Illustration of the composited data cube for each of the five AEZs. The 30-m Data-cube for South Asia composited three time-periods using 2013–2015 Landsat-8 Operational Land Imager (OLI) data. For each period (e.g. **Period 1**, Monsoon season: Julian days 151–300 of a year; **Period 2**, winter or rabi season: 301–365 days of a years plus 1 to 90 days of next year; **Period 3**, summer season: Julian days 91–150 days), 10 bands (blue, green, red, NIR, SWIR1, SWIR2, TIR1, EVI, NDVI, NDWI) per period plus slope band were composited and a median value was calculated for a given pixel over the period. From the three periods, there was a 31-band mega-file data cube in GEE image collection.

- Ground data (Figure 3 shows the distribution of reference training data but not the validation data) obtained through field visits; and
- Sub-meter to 5-m very high-resolution Imagery (VHRI) data (Figure 4 shows the distribution of reference training data and not validation data)

The reference training and validation data from ground surveys were gathered during various field campaigns that precisely identified latitude and longitude using geographic position systems (GPS) systems. The reference training and validation data were also gathered from sub-meter to 5-m very high-spatial resolution imagery (VHRI) which come from numerous satellites such as IKONOS, Quickbird, Geoeye, and Worldview obtained through sub-meter to 5-m imagery sourced through the National Geospatial Agency (NGA). Both the ground data and VHRI data were gathered from 3×3 pixel homogeneous locations to obtain pure samples for training and validation. A wide array of variability in these samples were compiled and categorized, for example various crop types (e.g., whether they were irrigated (Figure 5) or rainfed (Figure 6) or croplands as seen

in VHRI (Figure 7). Capturing a diverse set of data during various time-periods allow us to capture the diversity of agriculture whether they are irrigated (Figure 5) or rainfed (Figure 6) or whether they are croplands or non-croplands (Figure 3 through 7).

3.2.1. Reference data organization for training and validation

Reference data is obtained from ground surveys as well as sub-meter to 5 m very high-resolution imagery (VHRI) (see section 3.2 through 7). From these both sources we obtained a total of 2179 reference training samples for training the Random Forest (RF) machine learning algorithms (MLAs). Of the 2179 samples, 986 samples were cropland samples and the remaining 1193 samples were non-cropland samples (Table 3). The spatial distribution of these reference samples is shown in Figures 3 and 4. For validation (section 3.2.2) we have a total of 1185 samples of which 917 came from ground data and the 268 came from VHRI. Location of these is not shown in Figures 3 or 4. As shown in Table 3, the number of samples, for both training and validation, were much

Table 3. Number of reference training samples of croplands versus non-croplands in each of the five refined agro-ecological zones (AEZs) in South Asia.

RAEZ #	RAEZ Climate	Cropland training samples #	Non-cropland training samples #	Total training samples #	Cropland validation samples	Non-cropland validation samples	Total validation samples	Land area Km ²	Area per sample Km ² /#
1	Arid	21	24	45	84	39	123	228,140	5070
2	Semi-Arid	144	87	231	142	83	225	939,845	4069
3	Sub humid	392	507	899	315	35	350	1,455,099	1619
4	Tropics Humid	346	457	803	238	56	294	1,180,923	1471
5	Tropics Multiple Climates	83	118	201	138	55	193	548,697	2730
	Total	986	1193	2179	917	268	1185	435,2704	14957

RAEZ = Refined agro-ecological zones. The RAEZs were derived based on first taking FAO's global agro-ecological zones and refining them by taking local agricultural conditions and then aggregating them to five RAEZs.

higher in Semi-arid, sub-humid, and humid zones because they, together, cover 98% of all cropland areas of South Asia. The arid (Table 3) have only about 2% of the cropland areas of south Asia and hence have proportionally much lower training and validation samples.

3.2.2. Validation data

The total number of reference training and validation data were 3364 samples (2179 for training and 1185 for validation; Table 3). So a total of 35% of the samples were reserved for validation. Validation was conducted separately for each of the five AEZs (Figure 1). Validation involved developing error matrices that provide, for each of the five AEZs as well as for the entire area overall accuracies, errors of omissions, and errors of commissions. The entire sample training and validation dataset used is made available at the following website: <https://croplands.org/app/data/search>.

4. Methods

An overview of the proposed methods and approaches is shown in Figure 8. The analysis starts with an imagery stack or mega file data cube (MFDC) as shown below in Figure 2. Such MFDC's were composed for each of the five agro-ecological zones (AEZ's) of South Asia (Figure 1). The 31 band MFDC (Figure 2) were generated separately for each of the five AEZs. A custom cloud detecting script which was available to us from our team member (Oliphant et al. 2017a; 2017b) was used to run within GEE to mask out cloud pixels. The novel script ran faster and masked out more cloud impacted pixels efficiently than Fmask (Housman et al. 2015; Zhu et al. 2015). All the image processing steps and big-data computing for large AEZ areas were performed on the Google Earth Engine (GEE) cloud computing platform to enable seamless and fast work.

4.1. Random forest (RF) machine learning algorithm

Using GEE, a pixel-based supervised random forest (RF) machine learning algorithm was used for classification using data cubes (e.g. Figure 2) for each of the five AEZs (Figure 1). Random forest classifiers applied to Landsat imagery in GEE have successfully mapped a variety of vegetation including plantations such as oil palms (Lee et al. 2016) and rubber (Beckschäfer 2017). The RF classifier is a relatively fast, nonlinear classifier that is robust to noisy data (Rodriguez-Galiano et al. 2012). It uses multiple decision trees to assign classification labels and to reduce overfitting, furthermore, each tree is created from a subsection of training data.

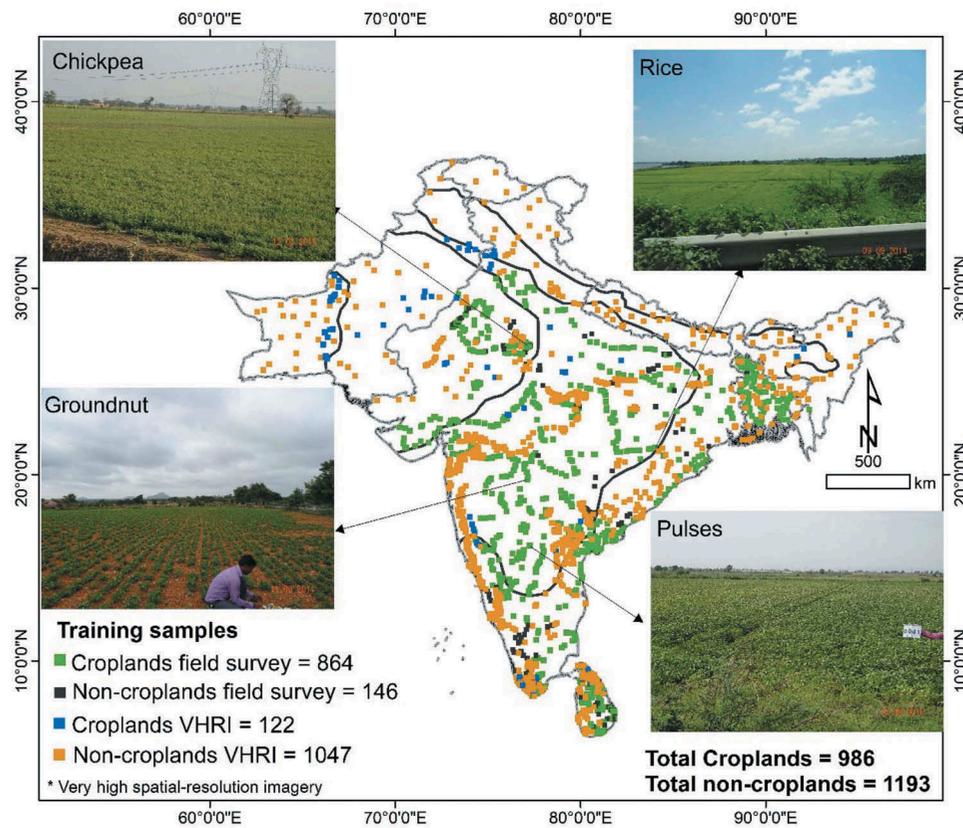


Figure 3. Reference training and validation data with illustrative ground data photographic samples of croplands versus non-croplands are shown in this figure. Location of the reference data collected by ground survey as well as sub-meter to 5 m very high-resolution imagery (VHRI) are shown on the map. Illustration of photographs taken during field visits of croplands versus non-croplands are also shown.

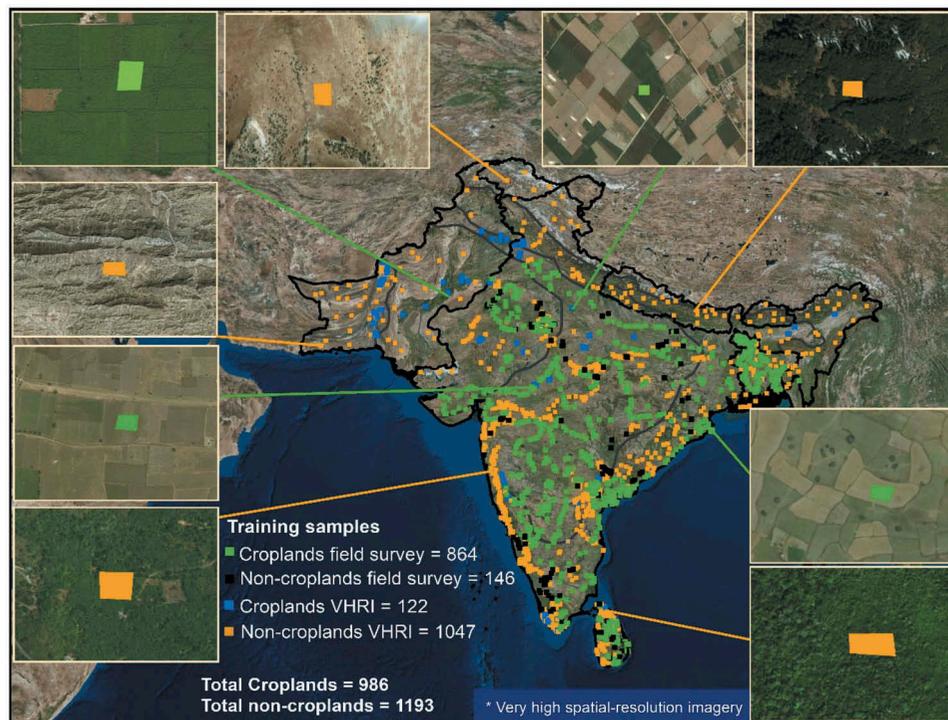


Figure 4. Reference training and validation data with sub-meter to 5-m VHRI illustrative samples of croplands versus non-croplands are shown in this figure. Location of the reference data collected by ground survey as well as sub-meter to 5 m very high-resolution imagery (VHRI) are shown on the map. Illustration of VHRI snapshots of croplands versus non-croplands are also shown.

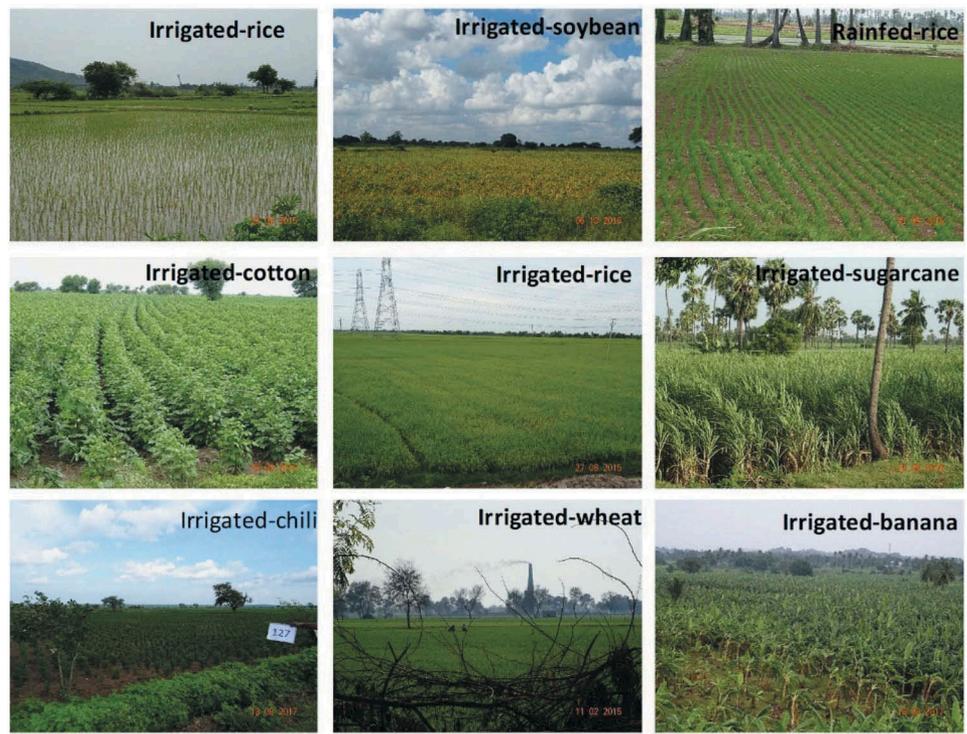


Figure 5. Illustrative ground-level photos of reference training and validation data samples for some irrigated areas.

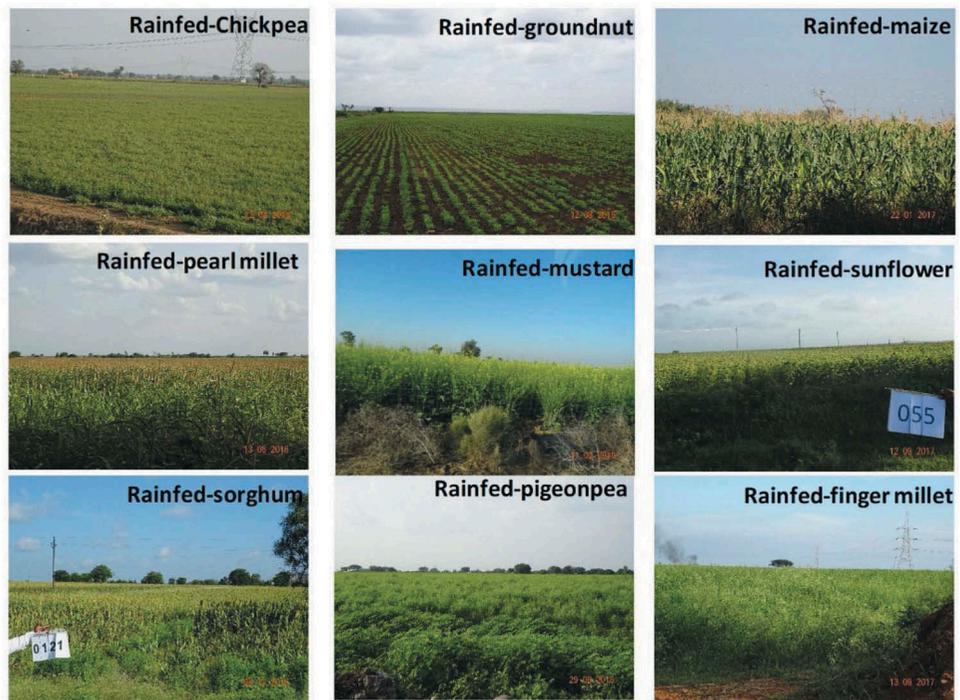


Figure 6. Illustrative ground-level photos of reference training and validation data samples for some rainfed areas.

Since the RF classification is a pixel-based supervised classifier, accurate cloud-free reference data were needed with a high-quality input raster. The classification of South Asia requires large reference data because of variability in spectral signatures of varied land use/

land cover (LULC) within this large geographical area. In this regard, RF classification for each of the five AEZ produced far better results than classifying the whole area, an iterative approach was used to add and remove training samples, based on classification results, to

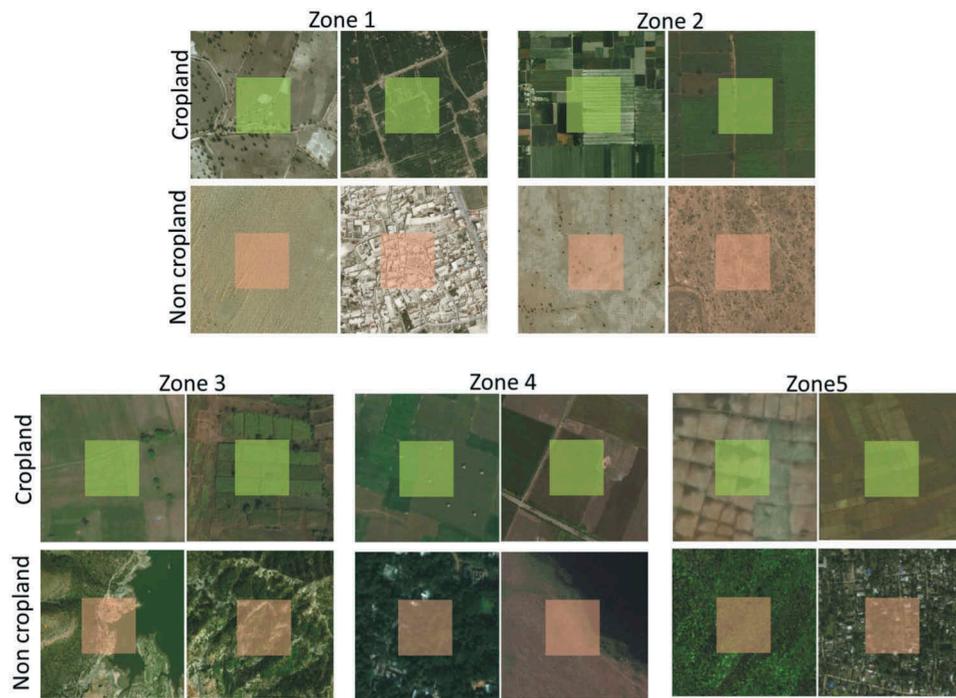


Figure 7. Illustrative reference sample selection from cropland versus non-cropland locations using sub-meter to 5-m very high-resolution imagery (VHRI).

improve the map classification and each map was visually assessed to see how well the classified map correlated with observed cropland in the landscape as seen in the sub-meter to 5-m VHRI (Oliphant et al. 2014) primarily using WorldView-3 data. The RF MLA was mainly used to generate cropland and non-cropland classes in five AEZs.

The sample size constantly increases with increasing iterations in RF classifier to map better classification in complex AEZs. The following steps (Figure 8) were considered invaluable:

- (1) Start with the best set of training samples for the RF classifier representing any one AEZ;
- (2) Extract the band values of the 31-band mosaic (e.g. Figure 2) of pixels that are in the training dataset to build an RF classification model knowledge-base (Figure 9) for each of the five AEZ's (Figure 1);
- (3) Use above knowledge-based (Figure 9) in RF classifier to classify the 30 m, 31-band MFDC data stack (Figure 2);
- (4) Visually interpret the classification results with independent ground data (not used in the classification or reserved for validation), existing reference maps from various sources (Teluguntla et al. 2015) as well as sub-meter to 5-m very high-spatial resolution imagery (VHRI);

- (5) If any misclassification, reclassify using additional or refined training samples with improved knowledge-base until the desired results of separating all or nearly all croplands from non-croplands are achieved. This is visually checked in detail using reference maps, sub-meter to 5-m VHRI;
- (6) Develop accuracy error matrices for AEZ and repeat the above steps until desired high levels of classification accuracies are achieved; and
- (7) Repeat the above steps for other AEZ's (Figure 1) completing the classification of the entire South Asia.

The number of iterations required to achieve better classification correlates to the complexity of the area as well as the quality of input datasets.

Based on prior experimentation and referencing works that systematically analyzed classification parameters (Heydari and Mountrakis 2018; Breiman 2001) we selected the following parameters for the random forest classification. 500 classification trees were chosen as it produced better classifications than fewer trees and had a reasonable processing time (≈ 1 h per zone). The default parameters were chosen for variables Per Split: $\sqrt{(n_bands)}$, minLeafPopulation: 1, and bag Fraction: 0.5. All computations were performed on GEE cloud.

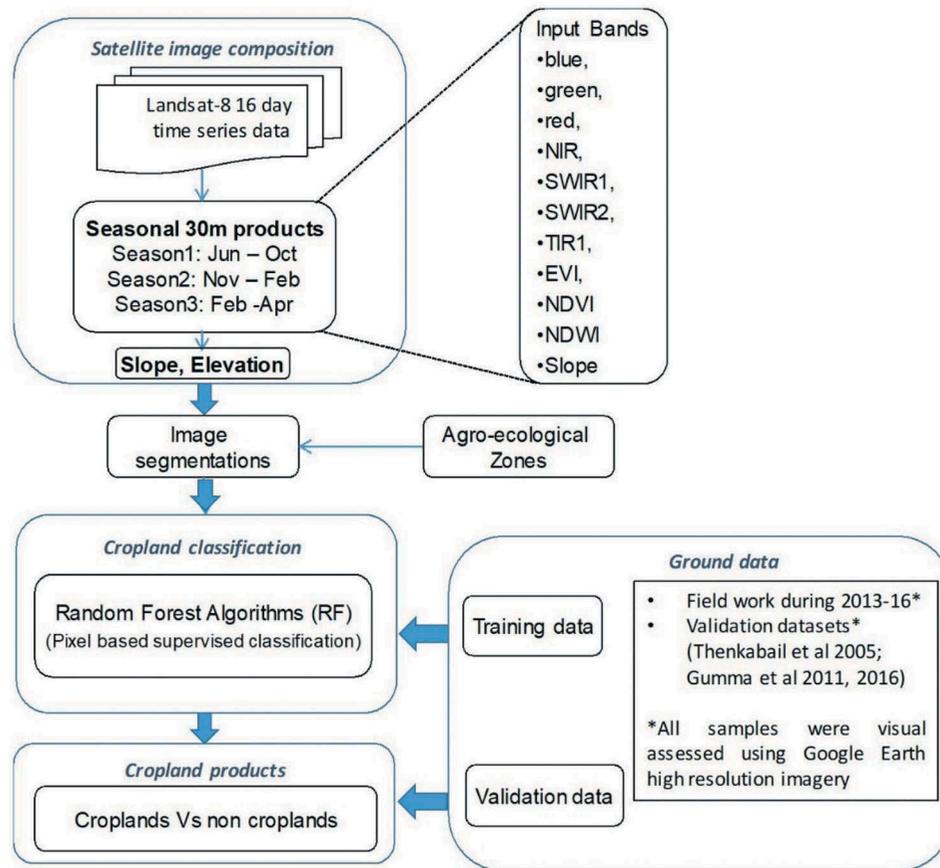


Figure 8. Methodology for mapping 30-m cropland extent. Pixel-based Random Forest (RF) machine learning algorithm for mapping cropland extent using Landsat 30-m data.

4.2. Cloud computing on Google Earth Engine

Due to large area classification, the size of the Landsat data is very high (running into peta-bytes when time-series is involved). It is very hard to process the data in regular systems, so we performed using Google Earth Engine cloud computing platform for image processing. GEE has the entire Landsat archive along with many openly available raster datasets from NASA, European Space Agency (ESA), and other imagery which allows code to be brought to data; complex multi-temporal continental scale data can be analyzed using relatively simple JavaScript or Python code and also can be shared and replicated by other researchers, lowering the barriers to utilizing super computers to perform geospatial analysis (Gorelick et al. 2017).

In this study, we used the GEE cloud-based supercomputing platform to integrate many capabilities that include:

- (1) Seamless Landsat image collection from 2014–2015 for the entire South Asia (Figure 2, Section 3.1);
- (2) Ability to write scripts in Python and JavaScript to process the petabyte scale data processing over very large areas such as South Asia within minutes using the GEE platform (Gorelick et al. 2017);

- (3) Machine learning algorithms (Section 4.1)
- (4) Ability for powerful and fast parallel computing linking 1000s of computers on the GEE cloud (Gorelick et al. 2017).

The entire classification process was performed in GEE using the above process. First, codes were executed on GEE to create cloud-free image composites. After creating the composites (Figure 2), the extracted training data's median pixel values were used as the knowledge-basis (Figure 9) for the pixel-based supervised RF MLA (Section 4.1) classification of each of the AEZ's of South Asia (Figure 1). After classification, the results were viewed within GEE, and the accuracy of classification was tested using reference data (e.g. independent field data from ground, data from collaborators, other available maps, and sub-meter to 5-m very high-resolution imagery or VHRI; Section 3.2). This process was iterated using additional training data until optimal classification results of croplands versus non-croplands were achieved.

The knowledge-base data (Figure 9) for the RF algorithm (Section 4.1) was calculated by overlaying reference training data (e.g. Table 3) on the mega file data cube (e.g. Figure 2). Next, the values were plotted in a box and

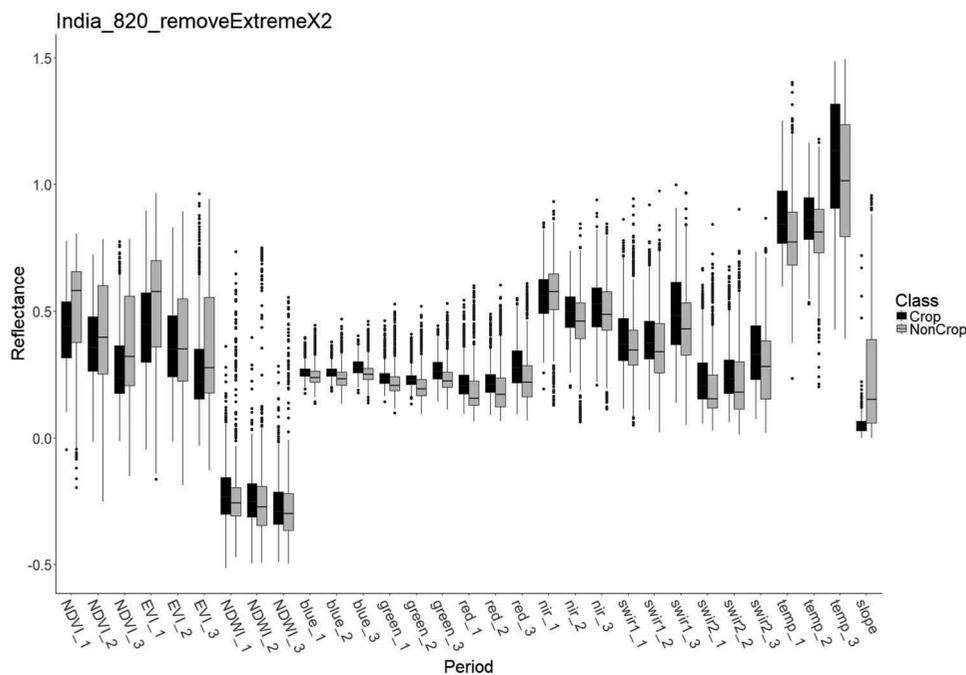


Figure 9. Knowledge generation for the random forest machine learning algorithm. Croplands versus non-croplands knowledge capture to train random forest machine learning algorithm using Landsat OLI data cube for each of the five agro-ecological zones (AEZs) of South Asia. Each data-cube consisted of [Band 1–3 (NDVI), band 4–6 (EVI), band 7–9 (NDWI), band 10–12 (blue), band 13–15 (green) 16–18 (red), 19–21 (NIR), 22–24 (SWIR1), 25–27 (SWIR2), 28–30 (TIR1), 31 (Slope)]. Note: TIR1 (thermal infrared band 1) values are scaled for convenience.

whisker plot in Figure 9 shows time-composited bands, one plot for each time-period (Table 2) and for three time periods (Monsoon, Summer, Winter), there are three time-composited bands. For example, bands 1 to 3 have three NDVI time-composited bands representing monsoon (DOY 151–300), summer (DOY 301–365 and DOY 1–60) and winter (DOY 61–150). Similarly, there are three EVI's bands for the same time-periods and so on. Overall, the 31 band MFDC (Figure 2) is stacked separately for each of the five AEZ.

The above approach of knowledge-generation, based RF algorithms is performed for each of the five agro-ecological zones (AEZs) of South Asia (Figure 1) and classified images into croplands versus non-croplands.

4.3. Classification post-processing

Once the final product of cropland versus non-croplands is achieved, post-processing is performed to sieve out “salt and pepper” noise that is inherent in any pixel-based classification product. A 3×3 median value kernel smoothing algorithm was used to sieve out noisy pixels and group non-contiguous pixels into contiguous pixels. However, fine features like roads and waterbodies will provide greater clarity to the product and hence were overlaid back on the

final product (Figure 10) using data from secondary sources. However, these features (roads, waterbodies) are part of non-croplands and hence are treated as such in accuracy assessments (Table 4).

4.4. Accuracy assessment

Accuracy assessment of the classification image was performed using validation data (Section 3.3, Table 3). Table 3 shows the distribution of the cropland versus non-cropland validation samples. These data are independent of the reference training data and reserved for validation purpose only and not used either in class classification or class identification. Accuracies were conducted using error matrices (Congalton and Kass 2008; Congalton et al. 2017) and established for each of the five AEZ's (Figure 1) as well as for the entire South Asia. Overall accuracies, producer's accuracies (or errors of omissions) and user's accuracies (or errors of commissions) were established.

4.5. Area calculations

Post-processing was also performed using GEE and cropland areas were calculated in GEE based on the Global Administrative Unit Layers (GAUL) database by the United

Nation's (UN) Food and Agricultural Organization or FAO (http://www.foodsec.org/tools_gaul.htm) country political boundaries. Since the 30-m cropland product (Figure 10) is of high-spatial resolution, areas are calculated at various geographic units. In this study, we calculated areas of countries as well as districts for entire South Asia and compared them with the area statistics available to us from the National statistics for these countries.

5. Results

This section summarizes the results of the Landsat data-derived 30-m cropland extent product of all South Asia (Figure 10). The results of accuracy assessment of the 30-m cropland product are also shown in Tables 4-6. The classified cropland versus non-cropland (Figure 10) derived areas are provided in Table 7. The 30-m product is compared with existing coarser resolution products (Figure 11). The cropland

areas derived from the 30-m products were compared with the national and sub-national crop area statistics obtained from the national systems of the South Asian countries (Figure 12). The product is made available on NASA's The Land Processes Distributed Active Archive Center (LP DAAC): <https://lpdaac.usgs.gov/products/gfsad30saafgircev001/>

5.1. Random forest classification results of croplands versus non-croplands

Methods described in section 4.0 and its sub-sections were applied using the pixel-based supervised random forest (RF) machine learning algorithm to establish croplands versus non-croplands for entire South Asia (Figure 10). This study created a 30-m seamless cropland extent product of South Asia (Figure 10) using 2014–2015 time-series Landsat-7 and Landsat-8 satellite data, and supervised random forest

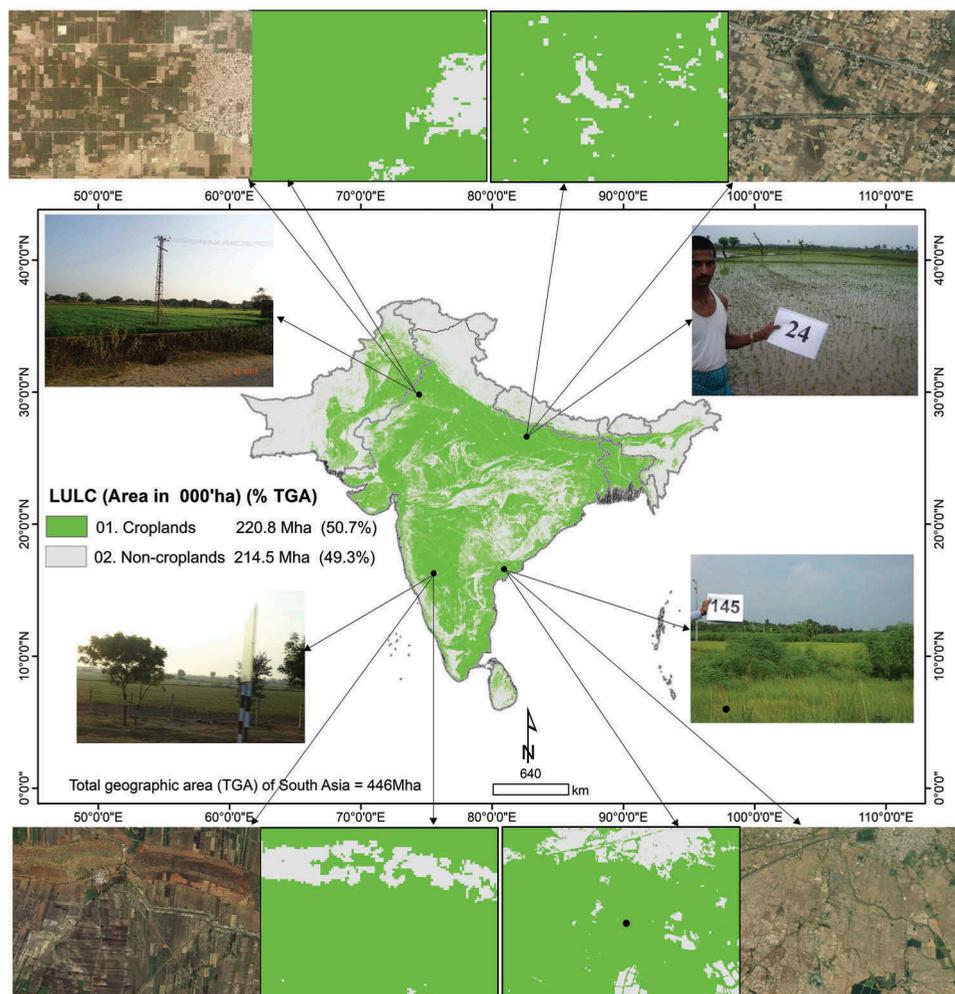


Figure 10. The South Asia 30-m cropland extent. The 30-m cropland extent derived using Landsat OLI data for the nominal year 2015. Zoom-in views highlight the detail provided for the 30-m product. For detailed zoom-ins visit www.croplands.org. The data can be downloaded from: <https://lpdaac.usgs.gov/products/gfsad30saafgircev001/>.

Table 4. Accuracy error matrix of the entire study area. Accuracy assessment error matrix of the 30-m cropland extent product.

Entire South Asia		reference data			User accuracy
		Crop	No-Crop	Total	
Map Data	Crop	824	41	865	95.3%
	No-Crop	93	227	320	70.9%
Total		917	268	1,185	
Producer accuracy		89.9%	84.7%		88.7%

Table 5. Error matrices of RAEZ 1, 2, 3, 4, and 5 of South Asia.

Zone 1		Reference Data			User Accuracy
		Crop	No-Crop	Total	
Map Data	Crop	75	6	81	92.6%
	No-Crop	9	33	42	78.6%
Total		84	39	123	
Producer Accuracy		89.3%	84.6%		87.8%

Zone 1: Geographic area 5.3% (total geographic area = 435.3 Mha)

Zone 1: Cropland area 0.1% (total net cropland area = 220.2 Mha)

Zone 2		Reference Data			User Accuracy
		Crop	No-Crop	Total	
Map Data	Crop	130	15	145	89.7%
	No-Crop	12	68	80	85.0%
Total		142	83	225	
Producer Accuracy		91.6%	81.9%		88.0%

Zone 2: Geographic area 22.27% (total geographic area = 435.3 Mha)

Zone 2: Cropland area 24% (total net cropland area = 220.2 Mha)

Zone 3		Reference Data			User Accuracy
		Crop	No-Crop	Total	
Map Data	Crop	276	6	282	97.9%
	No-Crop	39	29	68	42.7%
Total		315	35	350	
Producer Accuracy		87.6%	82.9%		87.1%

Zone 3: Geographic area 33.01% (total geographic area = 435.3 Mha)

Zone 3: Cropland area 45.6% (total net cropland area = 220.2 Mha)

Zone 4		Reference Data			User Accuracy
		Crop	No-Crop	Total	
Map Data	Crop	223	4	227	98.2%
	No-Crop	15	52	67	77.6%
Total		238	56	294	
Producer Accuracy		93.7%	92.9%		93.5%

Zone 4: Geographic area 27.15% (total geographic area = 435.3 Mha)

Zone 4: Cropland area 28.4% (total net cropland area = 220.2 Mha)

Zone 5		Reference Data			User Accuracy
		Crop	No-Crop	Total	
Map Data	Crop	120	10	130	92.3%
	No-Crop	18	45	63	71.4%
Total		138	55	193	
Producer Accuracy		86.9%	81.8%		85.5%

Zone 5: Geographic area 12.27% (total geographic area = 435.3 Mha)

Zone 5: Cropland area 1.9% (total net cropland area = 220.2 Mha)

machine learning algorithms, trained using extensive reference cropland and non-cropland data. The results of this product is discussed below.

Table 6. 30-m Landsat derived Cropland areas by country compared with national statistics.

Country	01. This study: GFSAD30 Croplands (Mha)	02. This study: GFSAD30 Non croplands (Mha)	03. This study: Total croplands plus non-croplands (Mha)	04. National statistics: Croplands (Mha)
India	184.3	142.2	326.4	168.1
Pakistan	22.4	57.7	80.1	22.9
Bangladesh	9.97	4.1	14.1	9.8
Nepal	2.19	12.6	14.8	4.03
Sri Lanka	1.84	4.7	6.6	2.08
Bhutan	0.06	3.9	4.0	0.12
Total	220.8	225.3	446.0	207.0

Sources:

<http://www.indiastat.com>.

<http://www.fao.org/faostat/>.

5.1.1. Croplands versus non-croplands knowledge capture

The South Asia cropland extent product is shown in Figure 10. In Figure 10, the top-left image “zoom-in” shows the identification of settlements among croplands whereas bottom-left “zoom-in” image identified barren land among croplands. The top-right “zoom-in” image shows the identification of forest, wastelands among the croplands whereas bottom-right “zoom-in” image identified settlements and waterbodies among croplands. The product shows croplands at a fine (1 pixel = 0.09 Ha) spatial resolution over the entire South Asia in great detail. The product is called the Global Food Security-support Analysis Data @ 30-m Cropland Extent for South Asia. A user can “zoom-in” to any part of South Asia and view croplands versus non-croplands at full resolution (www.croplands.org) and/or downloaded the product from NASA's LP DAAC:

<https://lpdaac.usgs.gov/products/gfsad30saafgirceV001/>

In Figure 10, cropland extent is shown in bright green and non-cropland extent is shown in gray. The benefit of the 30-m cropland maps is more visible when compared with existing coarse resolution cropland maps as shown from left to right in Figure 11: 30-m GFSAD30 (this study), 500 m GIAM (Dheeravath et al. 2010), 250-m Global Crop Extent (Pittman et al. 2010). These are compared with the reference sub-meter to 5-m VHRI. “Zoom-in” views of the four sub-areas are also shown along with the sub-meter to 5-m very high-spatial resolution imagery (VHRI) data corresponding to these “zoom-in” views for visual comparison (Figure 10). The 30-m cropland product provides detail and precision on where croplands are geographically compared to the uncertainty involved



Figure 11. Spatial comparison of cropland products for precision. Comparison of cropland extent products across three different spatial resolutions (30-m, 250-m, and 500-m) with <5-m very high-resolution imagery.

in coarser resolution data (Figure 11). Spatial comparison of cropland products across three different spatial resolutions (30-m, 250-m, and 500-m) and the reference sub-meter to 5-m VHRI are shown in Figure 11. The Coarser the pixel, the more approximate the cropland locations. Further, areas calculated from these coarser products will also have greater uncertainty.

5.1.2. Accuracy assessments

Accuracy assessment error matrices are reported for the entire South Asia (Table 4) as well as for each of the five AEZ's (Table 5). A total of 1185 well-dispersed validation samples (Section 3.3) were available for assessing the accuracies of South Asia. The results showed that the South Asia cropland product (Figure 10) had a producer's accuracy of 89.9% (errors of omissions of 10.1%), user accuracy of 95.3% (errors of commission of 4.7%) and overall accuracy of 88.7%. This implies the product (Figure 10) misses 10.1% of croplands and adds 4.7% of non-croplands (Table 4). Errors of omissions and commissions are consistently improved by gathering good reference data to train the RF machine learning algorithm (MLA). Large number of samples by themselves does not ensure high accuracies. Most important factors for higher accuracies were the following:

- Purity of samples used in training RF MLA (e.g. avoiding edge pixels, mixed pixels)

- Diversity of sampling (e.g. capturing different cropland types);

- Spatially well distributed samples;

- Ability to tweak the RF MLA through multiple iterations till the desired results are achieved;

- Number of samples

When all of the above conditions are met, accuracies at certain stage reach optimal, meaning further addition of samples will not increase accuracies. The goal not to just achieve high overall accuracies, which can be deceptive in the sense the class that we desire may or may not have high accuracies just because overall accuracies are high. The goal in this study was to achieve very high producer's accuracies (or very low errors of omissions) by ensuring the user's accuracies (errors of commissions) low. Often, when we try to decrease errors of omissions (croplands mapped as croplands) the errors of commissions go up (non-croplands mapped as croplands). Optimal results as shown in Tables 4 and 5 were achieved by trial and error by tweaking the RF MLA through multiple iterations by adding and deleting reference samples and generating much purer separability in the knowledge base for the RF MLA. With low errors of omissions 10.1% and low errors of commission of 4.7% for entire South Asia (Table 4), we achieved an optimal result in capturing as much croplands as possible.

Accuracies are also reported (Table 5) for each of the five AEZ's (Figure 1) of the South Asia cropland product (Figure 10). As shown in these tables, only 2% of the total net cropland areas (TNCA) of 220.2 Mha are in zones 1 and 5 put together, whereas zones 1, 2, 3, and 4 have 98% of the croplands. As a result, the validation data for zones 1 and 5 are much smaller than those of zone 2, 3, and 4 (Table 5). Overall for the five zones, the producer's

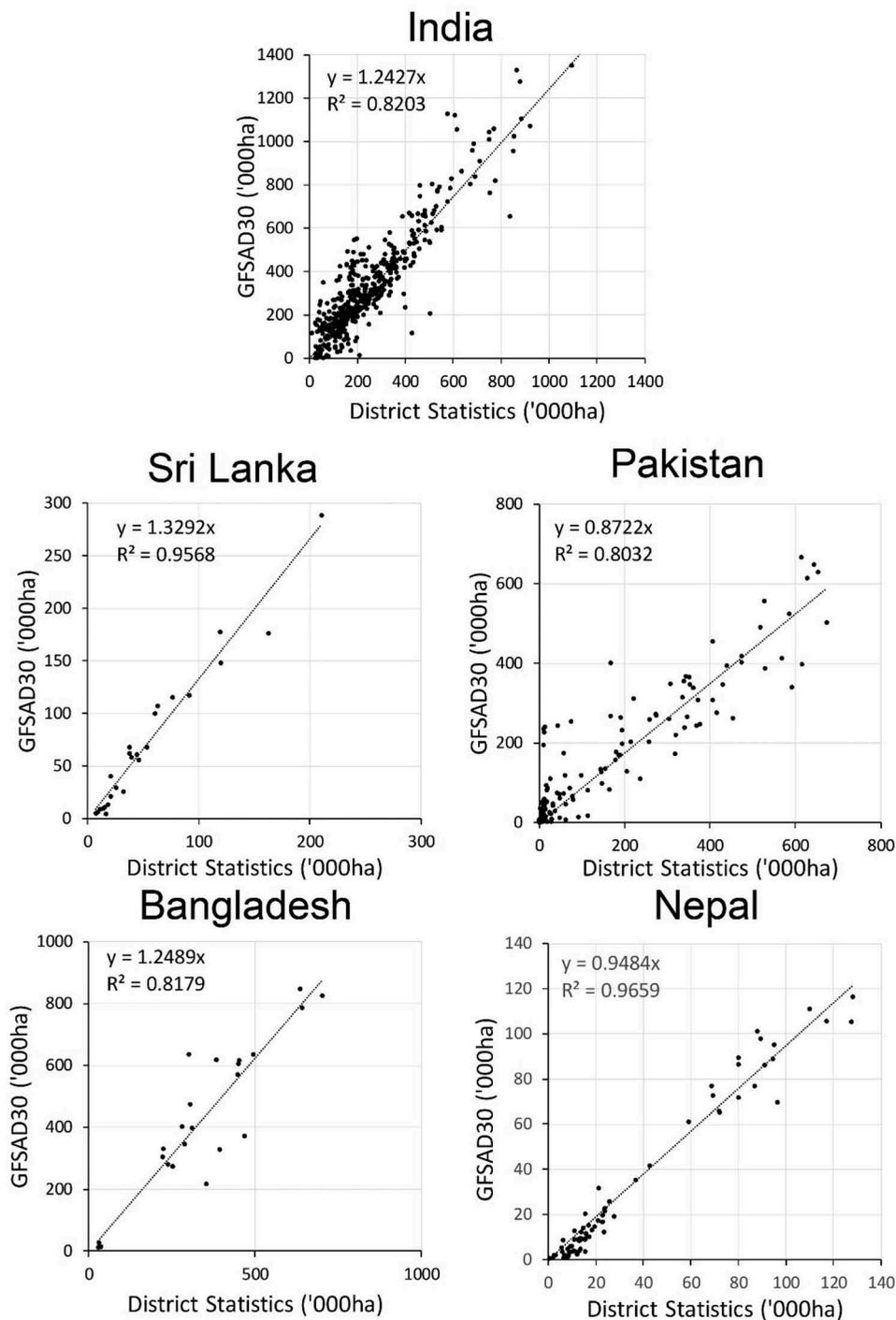


Figure 12. Comparison of district-wise cropland areas. 30-m Landsat derived versus national area statistics.

accuracies varied between 86.9% and 93.7% (errors of omissions varied from 6.3% to 13.1%), user's accuracies varied between 89.7% and 98.2% (errors of commissions varied from 1.8% to 11.3%), and overall accuracies varied between 85.5% and 93.5% (Table 5). All five zones achieved similar low errors of omissions and commissions (Table 5). Errors of commissions of 13.9% in zone 5 were mainly a result of complex mountain terrain.

5.2. Comparison landsat product with national statistical data

Table 6 compares cropland areas derived in this study (GFSAD30) from Landsat 30-m time-series data of 2014 and 2015 (called nominal 2015) with the cropland statistics obtained from the National systems of the South Asian Countries. Overall, there was 14.8 Mha more cropland areas estimated by GFSAD30 (this study) with 220.8

Mha compared to 207 Mha reported by the National systems as gathered from the UN FAO statistics (<http://www.fao.org/faostat/>); (Portmann, Siebert, and Döll 2010). These differences can be expected due to several reasons such as the type of data used, methodologies implemented and resolution of the imagery.

There was very a close match in cropland area estimates of Bangladesh and Pakistan (Table 6). Croplands of both these countries are relatively easy to assess due to long stretches of flatbed croplands along the river systems (e.g. Ganga and Brahmaputra in Bangladesh, Indus valley in Pakistan) and/or well-structured irrigated command areas. The greatest differences in cropland areas were in Nepal, Bhutan and Sri Lanka where remote sensing significantly underestimated National statistics. This was mainly a result of the complexity of croplands in the mountain terrains and approaches to estimating or mapping them. Most croplands of South Asia are in India with an area of 184.3 Mha as derived from GFSAD30 as opposed to 168.1 Mha estimated by the national statistics. India's croplands are spread across diverse agro-ecological zones, cropping practices, irrigation techniques, and rainfed areas. Further, croplands are either contiguous vast stretches in some places and discrete and fragmented in other places. They occur in all topographies: from flat flood plains and river courses to mountainous areas. Croplands in India also occur in various irrigation types (e.g. large reservoirs, small tanks, ground water pumping). High-resolution, time series remote sensing data from Landsat with 30-m spatial resolution offers the best opportunity to map all of these diverse croplands (Figure 10) which can be viewed in full resolution at www.croplands.org.

From the results in Table 7, it is clear that GFSAD30 derived cropland areas were comparable to MIRCA 2014 estimates (Portman and Siebert, personal communication) and GRIPC (Salmon et al. 2015). Since all these are either 2014 or 2015 estimates, this supports our findings. The FAO reported areas coming from the National systems were about 10% less than GFSAD30. The GIAM and GMRCA areas (Biradar et al. 2009; Thenkabail et al. 2009b) were for the nominal year 2010 based on much coarser resolution (1 to 10 km) remote sensing data of 2000–2010, therefore, significant lower areas of GIAM and GMRCA are expected (Table 7).

However, a much better comparison of cropland areas identified from the Landsat-derived 30-m product were included with sub-national statistics (Figure 12). For this district or equivalent administrative region, the statistics derived from this study were compared with the equivalent National statistics obtained through National systems of the corresponding countries (Figure 13). The results of this study explained 82% to

96% of the variability of district cropland areas obtained from the National statistics (Figure 12), specific countries included; India, Sri Lanka, Pakistan, Bangladesh, and Nepal R^2 and variances between the two datasets were 0.8203, 0.9568, 0.8032, 0.8179 and 0.9659, respectively.

6. Discussion

This study mapped cropland extent over the entire South Asia at the high-spatial resolution of 30-m utilizing Landsat time-series data of 2014–2015. This is an advancement over previous studies that mapped cropland extent in South Asia using coarser resolution data like MODIS at 250–1000 m resolution. In cropland mapping and studies pertaining to agricultural croplands, the need for near continuous time-series data are of significant importance to capture the phenology and growth characteristics of crops. Historically, the limitation in utilizing high resolution (30-m or better) imagery has been the unavailability of time-series data, especially over large areas. However, by using the availability of Landsat-8 and Landsat-7 images (every 8 days from the two satellites) this limitation has been reduced. When such 8-day data is acquired over multiple years (e.g. 2014 and 2015 as in this study) it further enhanced cloud-free imagery. In order to increase the temporal availability of high-resolution imagery, several studies (e.g. Oliphant et al. 2019; Teluguntla et al. 2018; Xiong et al. 2017a) have used data from multiple imaging platforms (e.g. Landsat and Sentinel) to produce cropland maps over large areas. Xiong et al. (2017a, 2017b) noted that the use of Landsat 8 along with Sentinel 2 temporal acquisition outweighs any minor drawbacks from mismatch in radiometry and geometry using multiple sensors. Nevertheless, a careful evaluation and synthesis of data from multiple platforms will harmonize and standardize any minor geometric and radiometric mismatch (Claverie et al. 2017). In this study, we increased the temporal frequency using Landsat 8 and 7, so there were no issues involved in the use of data from multiple platforms. Increasing temporal acquisition from multiple satellites also increases the performance of data interpolation and curve fitting algorithms (Brooks et al. 2012; Nguyen et al. 2012; Vuolo, Ng, and Atzberger 2017). In addition to Landsat 8, Landsat 9, is currently being constructed and is planned to be launched in 2020, ensuring the continuity and relevance of Landsat imagery (Jenstrom and Sauer 2018). When data from multiple Landsat systems become available (e.g. Landsat 9 along with Landsat 8 and 7), then combined with data from systems such as Sentinels such as Sentinel 2 we will have access to rich time-series data to conduct any sort of agricultural cropland studies.

Table 7. Country wise cropland areas compared between this study and other coarse resolution studies. Net cropland areas derived based on 30-m Global Food Security Support Analysis Data (GFSAD30) cropland product and comparison with other cropland products.

Country name	Land Area (Mha)	GFSAD30 ¹ (Mha)	MIRCA ² (Mha)	FAO ³ (Mha)	GRIPC ⁴ (Mha)	GIAM -GMRCA ⁵ (Mha)
India	326.4	184.3	177.4	169.7	187.5	150.1
Pakistan	80.1	22.4	25.2	21.3	20.4	17.7
Bangladesh	14.1	10.0	10.0	8.5	9.2	7.8
Nepal	14.8	2.2	3.4	2.5	3.5	4.4
Sri Lanka	6.6	1.8	2.2	2.2	3.5	2.4
Bhutan	4.0	0.1	0.2	0.1	0.1	0.2
Total	446.0	220.8	218.4	204.3	224.2	182.6

1 = GFSAD30 current study

2 = Monthly irrigated and rainfed crop areas (MIRCA) around the year 2014 derived by Portmann et al., 2010

3 = FAO Agricultural land area excluding pasture based on FAO2013 statistics consider nominal 2015 <http://www.fao.org/faostat/en/#data/QC>

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

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.

In this study, we first generated cloud-free image composites from Landsat 8 but through initial research, it was found that the composite of both Landsat 8 and Landsat 7 was far better than composite of Landsat 8 alone in providing seamless wall-to-wall coverage of cloud-free imagery over entire South Asia. In order maximize information and ability to separate croplands from non-croplands (Oliphant et al. 2019; Teluguntla et al. 2018; Xiong et al. 2017a, 2017b), all available spectral bands (blue, green, red, NIR, SWIR1, SWIR2, TIR) both in Landsat-8 and Landsat 7 along with three derived indices (NDVI, EVI, and NDWI) were included. These 10 bands were composited for each season (Monsoon, winter, and summer) and GDEM-derived slope data was added. The ability to separate croplands of varying kinds from non-croplands with a high degree of accuracies became possible as a result of such rich time-composited cloud-free imagery from multiple bands over all three seasons. In addition, Figure 9, shows the standard deviation of the bands, indices and slope layer that helped class separability significantly.

The random forest algorithm was used for the image classification because it has been successfully used for over a decade in remote landscape classifications, and is resistant to highly correlated data and for overcoming data over-fitting (Belgiu and Drăguț 2016). Because of this, high band correlation was not a factor for decreasing classifier performance. The continuous iterative process of classification, review, adding additional training samples of misclassified areas and reclassification was performed to improve the product. An independent validation data was used in the accuracy assessments and it is evident that accuracies are robust and reliable, especially considering this study covers large areas with heavy cloud cover especially during the monsoon season.

Comparing Landsat derived cropland areas (GFSAD30) with National statistics showed satisfactory results. In India, Landsat derived cropland area was

about 184.3 Mha compared to 168.1 Mha national statistics. Whereas in Pakistan it was 22.4 out of 22.9 Mha, Bangladesh it was 22.4 out of 22.9, Nepal it was 2.2 out of 4.0 Mha, Sri Lanka it was 1.8 out of 2.1 Mha, Bhutan it was 0.1 out of 0.1 Mha. Overall, the croplands of the above countries shows the variance of 220.8 Mha out of 207 Mha. The area statistics clearly reveal that an objective remote sensing-based cropland area estimates is feasible but methods and approaches can improve in high mountain areas.

In South Asia, the average farm size is meager 1.2 Ha, but as low as 0.3 Ha in Bangladesh (Lowder, Skoet, and Raney 2016; Otsuka 2014). However, there are three factors that help in overcoming the limitation of small farm size in mapping. These three factors are 1. Landsat pixel of 0.09 Ha (30 m x 30 m), 2. Contiguous nature of small farms, and 3. Methodology that involves the use of 8-day multi-year Landsat data. As a result, it was possible to capture even the smallest of farms, except when they are extremely fragmented and located in forests as “slash and burn” shifting cultivation or as extreme small fragments in other landscapes. Given extreme low percent (<1% of the total net cropland area) of such cases, this is not an issue.

In any classification, the goal is to obtain accuracies as high as possible and when producer’s accuracies (errors of omissions) increase, user’s accuracies (errors of commissions) decrease and vice versa. The random forest algorithm is iterated to achieve optimal producer’s and user’s accuracies. In order to serve as a foundation for all higher-level cropland products such as cropping intensity, irrigation or rainfed areas mapping and crop type mapping. It was preferable to improve the classification for producer’s accuracies (e.g. get all the croplands mapped as perfectly as possible even when user’s accuracies reduce slightly in trade-off). In other words, over-classifying cropland (non-cropland as cropland) was preferable to under-classifying cropland (cropland as non-cropland).

7. Conclusions

The study produced the high-spatial resolution Landsat-8 and 7 derived 30-m (1 pixel = 0.09 ha) cropland extent product of the nominal year 2015 (based on every 8 day data for the 2013–2015 time period) for the six of the 8 countries of South Asia consisting of India, Pakistan, Bangladesh, Nepal, Sri Lanka and Bhutan (Afghanistan and Maldives not included in the study) covering a geographic area of 435.3 Mha. Agricultural croplands of the entire South Asia was 220.8 Mha (50.7% of the geographic area). The product showed that the South Asia cropland product had a producer's accuracy of 89.9% (errors of omissions of 10.1%), user's accuracy of 95.3% (errors of commission of 4.7%) and overall accuracy of 88.7%. The National and sub-national (districts) areas computed from this product explained 80–96% variability when compared with the National statistics of the South Asian Countries obtained from the National system. Overall, for the five agro-ecological zones of South Asia, the producer's accuracies varied between 86.9% and 93.7% (errors of omissions varied from 6.3% to 13.1%), user's accuracies varied between 89.7% and 98.2% (errors of commissions varied from 1.8% to 11.3%), and overall accuracies varied between 85.5% and 93.5%. Agricultural cropland areas of the six countries were: 184.3 Mha for India, 22.4 Mha for Pakistan, 9.97 Mha for Bangladesh, 2.19 Mha for Nepal, 1.84 Mha for Sri Lanka, and 0.06 Mha for Bhutan. The National and sub-national (districts) areas computed from this product explained 80–96% variability when compared with the National statistics of the South Asian Countries.

The study established methods and approaches for peta-byte scale cloud computing on the Google Earth Engine (GEE) cloud platform using the Random Forest (RF) Machine Learning Algorithms (MLAs) which were trained using rich set of reference training data gathered from the field as well as sub-meter to 5-m very high-spatial resolution imagery (VHRI). An independent reference validation data was used to assess accuracies based on error matrices for each of the five AEZs as well as for the entire South Asia. The full-resolution imagery can be viewed at www.croplands.org and the data downloaded from the NASA and the USGS The Land Processes Distributed Active Archive Center (LP DAAC): <https://lpdaac.usgs.gov/products/gfsad30saafgircev001/>

Highlights

- Highest spatial resolution (30-m) cropland product covering entire South Asia for 2015;
- Unique cropland mapping methods involving petabyte-scale computing on cloud;

- Random forest machine learning algorithm applied over very large area;
- Overall accuracy of 88.7% with 89.9% of agricultural croplands captured;
- Public availability of 30-m cropland product, codes, and reference data

Acknowledgements

This research was supported by the CGIAR Research Program Climate Change, Agriculture and Food Security (CCAFA), the CGIAR Research Program Water, Land and Ecosystems (WLE) which are carried out with support from the CGIAR Trust Fund and through bilateral funding agreements. For details visit <https://ccafs.cgiar.org/donors> and <https://wle.cgiar.org/donors>. Funding was also from NASA MEaSUREs, through the NASA ROSES solicitation, for a period of 5 years (1 June 2013–31 May 2018). The NASA Making Earth System Data Records for Use in Research Environments (MEaSUREs) grant number is NNH13AV82I and the USGS Sales Order number is 29039. We gratefully acknowledge this support. The United States Geological Survey (USGS) provided significant direct and indirect supplemental funding through its Land Resources Mission Area (LRMA), National Land Imaging Program (NLIP), and Land Change Science (LCS) program. We gratefully acknowledge this. This research is a part of the Global Food Security – support Analysis Data Project at 30-m (GFSAD30). The project was led by USGS in collaboration with NASA Ames, University of New Hampshire (UNH), University of Wisconsin (UW), NASA Goddard Space Flight Center (GSFC), and the Northern Arizona University (NAU). Finally, special thanks to Ms. Susan Benjamin, Director of Western Geographic Science Center (WGSC) of USGS, Mr. Larry Gaffney, Administrative officer of WGSC of USGS for their support and encouragement throughout this research. We would like to thank to Dr Sunil Dubey, Assistant director, MNCFC for providing sub-national statistics.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the CGIAR Research Program Water, Land and Ecosystems (WLE) [None]; NASA MEaSUREs, through the NASA ROSES solicitation, [NNH13AV82I]. The United States Geological Survey (USGS) provided significant direct and indirect supplemental funding through its Land Resources Mission Area (LRMA), National Land Imaging Program (NLIP), and Land Change Science (LCS) program.

ORCID

Murali Krishna Gumma  <http://orcid.org/0000-0002-3760-3935>

Prasad S. Thenkabail  <http://orcid.org/0000-0002-2182-8822>
Pardhasaradhi G. Teluguntla  <http://orcid.org/0000-0001-8060-9841>

Adam Oliphant  <http://orcid.org/0000-0001-8622-7932>

Jun Xiong  <http://orcid.org/0000-0002-2320-0780>
 Chandra Giri  <http://orcid.org/0000-0003-1938-4131>
 Vineetha Pyla  <http://orcid.org/0000-0002-9436-0394>
 Sreenath Dixit  <http://orcid.org/0000-0002-3607-8729>
 Anthony M Whitbread  <http://orcid.org/0000-0003-4840-7670>

References

- Beckschäfer, P. 2017. "Obtaining rubber plantation age information from very dense Landsat TM & ETM + time series data and pixel-based image compositing." *Remote Sensing of Environment* 196: 89–100. doi:10.1016/j.rse.2017.04.003.
- Belgiu, M., and L. Drăguț. 2016. "Random Forest in Remote Sensing: A Review of Applications and Future Directions." *ISPRS Journal of Photogrammetry and Remote Sensing* 114: 24–31. doi:10.1016/j.isprsjprs.2016.01.011.
- Biggs, T. W., P. S. Thenkabail, M. K. Gumma, C. A. Scott, G. R. Parthasaradhi, and H. N. Turrall. 2006. "Irrigated Area Mapping in Heterogeneous Landscapes with MODIS Time Series, Ground Truth and Census Data, Krishna Basin, India." *International Journal of Remote Sensing* 27: 4245–4266. doi:10.1080/01431160600851801.
- Biradar, C. M., P. S. Thenkabail, P. Noojipady, Y. Li, V. Dheeravath, H. Turrall, M. Velpuri, M. K. Gumma, O. R. P. Gangalakunta, and X. L. Cai. 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. doi:10.1016/j.jag.2008.11.002.
- Breiman, L. 2001. "Random Forests." *Machine Learning* 45: 5–32. doi:10.1023/A:1010933404324.
- Brooks, E. B., V. A. Thomas, R. H. Wynne, and J. W. Coulston. 2012. "Fitting the Multitemporal Curve: A Fourier Series Approach to the Missing Data Problem in Remote Sensing Analysis." *IEEE Transactions on Geoscience and Remote Sensing* 50: 3340–3353. doi:10.1109/TGRS.2012.2183137.
- Claverie, M., J. G. Masek, J. Ju, and J. L. Dungan. 2017. *Harmonized Landsat-8 Sentinel-2 (HLS) Product User's Guide*. Washington, DC, USA: National Aeronautics and Space Administration (NASA).
- Congalton, G. R., K. Yadav, K. McDonnell, J. Poehnelt, B. Stevens, K. M. Gumma, P. Teluguntla, and P. S. Thenkabail. 2017. "NASA Making Earth System Data Records for Use in Research Environments (MEaSUREs) Global Food Security-support Analysis Data (GFSAD) Cropland Extent 2015 Validation Global 30 m V001 [Data Set]." *NASA EOSDIS Land Processes DAAC*. doi:10.5067/MEaSUREs/GFSAD/GFSAD30VAL001.
- Congalton, R. G., and K. Green. 2008. "Assessing the Accuracy of Remotely Sensed Data: Principles and Practices." *CRC Press/Taylor & Francis*. doi:10.1201/9781420055139.
- Defries, R. S., M. C. Hansen, and J. R. G. Townshend. 2000. "Global Continuous Fields of Vegetation Characteristics: A Linear Mixture Model Applied to Multi-year 8 Km AVHRR Data." *International Journal of Remote Sensing* 21: 1389–1414. doi:10.1080/014311600210236.
- Dheeravath, V., P. S. Thenkabail, G. Chandrakantha, P. Noojipady, G. P. O. Reddy, C. M. Biradar, M. K. Gumma, and M. Velpuri. 2010. "Irrigated Areas of India Derived Using MODIS 500 M Time Series for the Years 2001–2003." *ISPRS Journal of Photogrammetry and Remote Sensing* 65: 42–59. doi:10.1016/j.isprsjprs.2009.08.004.
- Dong, J., X. Xiao, W. Kou, Y. Qin, G. Zhang, L. Li, C. Jin, Y. Zhou, J. Wang, and C. Biradar. 2015. "Tracking the Dynamics of Paddy Rice Planting Area in 1986–2010 through Time Series Landsat Images and Phenology-based Algorithms." *Remote Sensing of Environment* 160: 99–113. doi:10.1016/j.rse.2015.01.004.
- Duro, D. C., S. E. Franklin, and M. G. Dubé. 2012. "A Comparison of Pixel-based and Object-based Image Analysis with Selected Machine Learning Algorithms for the Classification of Agricultural Landscapes Using SPOT-5 HRG Imagery." *Remote Sensing of Environment* 118: 259–272. doi:10.1016/j.rse.2011.11.020.
- Farr, T. G., P. A. Rosen, E. Caro, R. Crippen, R. Duren, S. Hensley, M. Kobrick, M. Paller, E. Rodriguez, and L. Roth. 2007. "The Shuttle Radar Topography Mission." *Reviews of Geophysics* 45. doi:10.1029/2005RG000183.
- Friedl, M. A., and C. E. Brodley. 1997. "Decision Tree Classification of Land Cover from Remotely Sensed Data." *Remote Sensing of Environment* 61: 399–409. doi:10.1016/S0034-4257(97)00049-7.
- Fritz, S., and L. See. 2008. "Identifying and Quantifying Uncertainty and Spatial Disagreement in the Comparison of Global Land Cover for Different Applications." *Global Change Biology* 14: 1057–1075. doi:10.1111/j.1365-2486.2007.01519.x.
- Giri, C., Z. Zhu, and B. Reed. 2005. "A Comparative Analysis of the Global Land Cover 2000 and MODIS Land Cover Data Sets." *Remote Sensing of Environment* 94: 123–132. doi:10.1016/j.rse.2004.09.005.
- Gorelick, N., M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore. 2017. "Google Earth Engine: Planetary-scale Geospatial Analysis for Everyone." *Remote Sensing of Environment* 202: 18–27. doi:10.1016/j.rse.2017.06.031.
- Gray, J., M. Friedl, S. Frolking, N. Ramankutty, A. Nelson, and M. K. Gumma. 2014. "Mapping Asian Cropping Intensity with MODIS." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7: 3373–3379. doi:10.1109/JSTARS.2014.2344630.
- Gumma, M. K., N. Andrew, P. S. Thenkabail, and N. S. Amrendra. 2011a. "Mapping Rice Areas of South Asia Using MODIS Multitemporal Data." *Journal of Applied Remote Sensing* 5: 053547. doi:10.1117/1.3619838.
- Gumma, M. K., P. S. Thenkabail, A. Maunahan, S. Islam, and A. Nelson. 2014. "Mapping Seasonal Rice Cropland Extent and Area in the High Cropping Intensity Environment of Bangladesh Using MODIS 500m Data for the Year 2010." *ISPRS Journal of Photogrammetry and Remote Sensing* 91: 98–113. doi:10.1016/j.isprsjprs.2014.02.007.
- Gumma, M. K., P. S. Thenkabail, F. Hideto, A. Nelson, V. Dheeravath, D. Busia, and A. Rala. 2011b. "Mapping Irrigated Areas of Ghana Using Fusion of 30 M and 250 M Resolution Remote-Sensing Data." *Remote Sensing* 3: 816–835. doi:10.3390/rs3040816.
- Gumma, M. K., P. S. Thenkabail, I. V. Muralikrishna, M. N. Velpuri, P. T. Gangadhararao, V. Dheeravath, C. M. Biradar, S. Acharya Nalan, and A. Gaur. 2011c. "Changes in Agricultural Cropland Areas between a Water-surplus Year and a Water-deficit Year Impacting Food Security, Determined Using MODIS 250 M Time-series Data and Spectral Matching Techniques, in the Krishna River Basin (India)." *International Journal of Remote Sensing* 32: 3495–3520. doi:10.1080/01431161003749485.
- Gumma, M. K., P. S. Thenkabail, K. C. Deevi, I. A. Mohammed, P. Teluguntla, A. Oliphant, J. Xiong, T. Aye, and

- A. M. Whitbread. 2018. "Mapping Cropland Fallow Areas in Myanmar to Scale up Sustainable Intensification of Pulse Crops in the Farming System." *GIScience & Remote Sensing* 55: 926–949. doi:10.1080/15481603.2018.1482855.
- Gumma, M. K., P. S. Thenkabail, P. Teluguntla, M. N. Rao, I. A. Mohammed, and A. M. Whitbread. 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* 9: 981–1003. doi:10.1080/17538947.2016.1168489.
- HarvestChoice. 2009. "Agroecological Zones." Accessed 1 May 2010. <http://harvestchoice.org/production/biophysical/agroecology>
- Herold, M., P. Mayaux, C. Woodcock, A. Baccini, and C. Schmullius. 2008. "Some Challenges in Global Land Cover Mapping: An Assessment of Agreement and Accuracy in Existing 1 Km Datasets." *Remote Sensing of Environment* 112: 2538–2556. doi:10.1016/j.rse.2007.11.013.
- Heydari, S. S., and G. Mountrakis. 2018. "Effect of Classifier Selection, Reference Sample Size, Reference Class Distribution and Scene Heterogeneity in Per-pixel Classification Accuracy Using 26 Landsat Sites." *Remote Sensing of Environment* 204: 648–658. doi:10.1016/j.rse.2017.09.035.
- Housman, I., V. Tanpipat, T. Biswas, A. Clark, P. Stephen, P. Maus, and K. Megown. 2015. "Monitoring Forest Change in Southeast Asia: Case Studies for USAID Lowering Emissions in Asia's Forests" (No. RSAC-10108-RPT1). Salt Lake City.
- Jenstrom, D., and B. Sauer (2018). "2018. Landsat 9 Project Status." In *Landsat Science Team Meeting*. US Geological Survey, Sioux Falls, SD.
- Lary, D. J., A. H. Alavi, A. H. Gandomi, and A. L. Walker. 2016. "Machine Learning in Geosciences and Remote Sensing." *Geoscience Frontiers* 7: 3–10. doi:10.1016/j.gsf.2015.07.003.
- Lee, J. S. H., S. A. Wich, A. Widayati, and L. P. Koh. 2016. "Detecting industrial oil palm plantations on Landsat images with Google Earth Engine." *Remote Sensing . Appl. Soc. Environment*. 4: 219–224. doi:10.1016/j.rsase.2016.11.003.
- Lobell, D. B., G. P. Asner, J. I. Ortiz-Monasterio, and T. L. Benning. 2003. "Remote Sensing of Regional Crop Production in the Yaqui Valley, Mexico: Estimates and Uncertainties." *Agriculture, Ecosystems & Environment* 94: 205–220. doi:10.1016/S0167-8809(02)00021-X.
- Lowder, S. K., J. Skoet, and T. Raney. 2016. "The Number, Size, and Distribution of Farms, Smallholder Farms, and Family Farms Worldwide." *World Development* 87: 16–29. doi:10.1016/j.worlddev.2015.10.041.
- Nguyen, T. T. H., C. De Bie, A. Ali, E. Smaling, and T. H. Chu. 2012. "Mapping the Irrigated Rice Cropping Patterns of the Mekong Delta, Vietnam, through Hyper-temporal SPOT NDVI Image Analysis." *International Journal of Remote Sensing* 33: 415–434. doi:10.1080/01431161.2010.532826.
- Oliphant, A., P. S. Thenkabail, P. Teluguntla, J. Xiong, M. K. Gumma, R. Congalton, and K. Yadav. 2019. "Mapping Cropland Extent of Southeast and Northeast Asia Using Multi-year Time-series Landsat 30-m Data Using Random Forest Classifier on Google Earth Engine." *International Journal of Applied Earth Observation and Geoinformation* 81: 110–124. doi:10.1016/j.jag.2018.11.014.
- Oliphant, A. J., J. Li, R. H. Wynne, P. F. Donovan, and C. E. Zipper, 2014. "Identifying woody vegetation on coal surface mines using phenological indicators with multitemporal landsat imagery. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences.*" *ISPRS Archives*. 339–345. doi:10.5194/isprsarchives-XL-1-339-2014.
- Oliphant, A. J., R. H. Wynne, C. E. Zipper, W. M. Ford, P. F. Donovan, and J. Li. 2017b. "Autumn olive (*Elaeagnus umbellata*) presence and proliferation on former surface coal mines in Eastern USA." *Biological Invasions* 19: 179–195. doi:10.1007/s10530-016-1271-6.
- Oliphant, A. J., P. S. Thenkabail, P. Teluguntla, J. Xiong, R. G. Congalton, K. Yadav, R. Massey, M. K. Gumma, and C. Smith. 2017a. "NASA Making Earth System Data Records for Use in Research Environments (MEaSUREs) Global Food Security-support Analysis Data (GFSAD) @ 30-m for Southeast & Northeast Asia: Cropland Extent Product (GFSAD30SEACE)." [WWW Document]. NASA EOSDIS L. Process. DAAC. doi:10.5067/MEaSUREs/GFSAD/GFSAD30SEACE.001.
- Otsuka, K. 2014. "Viability of Small-Scale Farms in Asia." In *Handbook on Food Demand, Supply, Sustainability and Security*, Chapter 18. edited by R. Jha, R. Gaiha, and A. B. Deolalikar, 462–481. Cheltenham, UK: Edward Elgar Publishing.
- Otukei, J. R., and T. Blaschke. 2010. "Land Cover Change Assessment Using Decision Trees, Support Vector Machines and Maximum Likelihood Classification Algorithms." *International Journal of Applied Earth Observation and Geoinformation* 12: S27–S31. doi:10.1016/j.jag.2009.11.002.
- Pal, M., and P. M. Mather. 2003. "An Assessment of the Effectiveness of Decision Tree Methods for Land Cover Classification." *Remote Sensing of Environment* 86: 554–565. doi:10.1016/S0034-4257(03)00132-9.
- Papademetriou, M. K. 2000. "Rice Production in the Asia-Pacific Region: Issues and Perspectives." In *Bridging the Rice Yield Gap in the Asia-Pacific Region*, edited by M. K. Papademetriou, F. J. Dent, and E. M. Herath, 220, Bangkok Food and Agriculture Organization of the United Nations.
- Pittman, K., M. C. Hansen, I. Becker-Reshef, P. V. Potapov, and C. O. Justice. 2010. "Estimating Global Cropland Extent with Multi-year MODIS Data." *Remote Sensing* 2: 1844. doi:10.3390/rs2071844.
- Portmann, F. T., S. Siebert, and P. Döll. 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. doi:10.1029/2008GB003435.
- Radoux, J., C. Lamarche, E. Van Bogaert, S. Bontemps, C. Brockmann, and P. Defourny. 2014. "Automated Training Sample Extraction for Global Land Cover Mapping." *Remote Sensing* 6: 3965. doi:10.3390/rs6053965.
- Rodriguez-Galiano, V. F., B. Ghimire, J. Rogan, M. Chica-Olmo, and J. P. Rigol-Sanchez. 2012. "An assessment of the effectiveness of a random forest classifier for land-cover classification." *ISPRS Journal Photogrammetry and Remote Sensing* 67: 93–104. doi:10.1016/j.isprsjprs.2011.11.002.
- Roy, P. S., M. D. Behera, M. S. R. Murthy, A. Roy, S. Singh, S. P. S. Kushwaha, C. S. Jha, et al. 2015. "New Vegetation Type Map of India Prepared Using Satellite Remote Sensing: Comparison with Global Vegetation Maps and Utilities." *International Journal of Applied Earth Observation and Geoinformation* 39: 142–159. doi:10.1016/j.jag.2015.03.003.
- Salmon, J. M., M. A. Friedl, S. Froking, D. Wisser, and E. M. Douglas. 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. doi:10.1016/j.jag.2015.01.014.
- Sangeeta, S., J. Gensuo, and Z. Anzhi. 2018. "Satellite View of Seasonal Greenness Trends and Controls in South Asia." *Environmental Research Letters* 13: 034026. doi:10.1088/1748-9326/aaa866.
- Siebert, S., P. Döll, J. Hoogeveen, J.-M. Faures, K. Frenken, and S. Feick. 2005. "Development and Validation of the Global Map of Irrigation Areas." *Hydrology and Earth System Sciences Discussions* 2: 1299–1327. doi:10.5194/hessd-2-1299-2005.
- Teluguntla, P., P. Thenkabail, J. Xiong, M. Gumma, C. Giri, C. Milesi, M. Ozdogan, R. Congalton, and J. Tilton. 2015. "Global Food Security Support Analysis Data (GFSAD) at Nominal 1 Km (GCAD) Derived from Remote Sensing in Support of Food Security in the Twenty-first Century: Current Achievements and Future Possibilities." Chapter 6. In *Remote Sensing handbook Volume II: Land Resources: Monitoring, Modeling, and Mapping*, edited by P. Thenkabail, 131–159. Boca Raton, FL: Taylor and Francis/CRC Press.
- Teluguntla, P., P. S. Thenkabail, A. Oliphant, J. Xiong, M. K. Gumma, R. G. Congalton, K. Yadav, and A. Huete. 2018. "A 30-m Landsat-derived Cropland Extent Product of Australia and China Using Random Forest Machine Learning Algorithm on Google Earth Engine Cloud Computing Platform." *ISPRS Journal of Photogrammetry and Remote Sensing* 144: 325–340. doi:10.1016/j.isprsjprs.2018.07.017.
- Teluguntla, P., P. S. Thenkabail, J. Xiong, M. K. Gumma, R. G. Congalton, A. Oliphant, J. Poehnelt, K. Yadav, M. Rao, and R. Massey. 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* 10-2017 (9). IP-074181. doi:10.1080/17538947.2016.1267269.
- Thenkabail, P., P. GangadharaRao, T. Biggs, M. Gumma, and H. Turrall. 2007a. "Spectral Matching Techniques to Determine Historical land-use/land-cover (LULC) and Irrigated Areas Using Time-series 0.1-degree AVHRR Pathfinder Datasets." *Photogrammetric Engineering & Remote Sensing* 73: 1029–1040.
- Thenkabail, P. S., C. M. Biradar, P. Noojipady, V. Dheeravath, Y. Li, M. Velpuri, M. Gumma, O. R. P. Gangalakunta, H. Turrall, and X. Cai. 2009a. "Global Irrigated Area Map (GIAM), Derived from Remote Sensing, for the End of the Last Millennium." *International Journal of Remote Sensing* 30: 3679–3733. doi:10.1080/01431160802698919.
- Thenkabail, P. S., J. W. Knox, M. Ozdogan, M. K. Gumma, R. G. Congalton, Z. Wu, C. Milesi, A. Finkral, M. Marshall, and I. Mariotto. 2012. "Assessing Future Risks to Agricultural Productivity, Water Resources and Food Security: How Can Remote Sensing Help?" *Photogrammetric Engineering and Remote Sensing* 78: 773–782.
- Thenkabail, P. S., M. Schull, and H. Turrall. 2005. "Ganges and Indus River Basin Land use/land Cover (LULC) and Irrigated Area Mapping Using Continuous Streams of MODIS Data." *Remote Sensing of Environment* 95: 317–341. doi:10.1016/j.rse.2004.12.018.
- Thenkabail, P. S., P. GangadharaRao, T. Biggs, M. K. Gumma, and H. Turrall. 2007b. "Spectral Matching Techniques to Determine Historical Land use/Land Cover (LULC) and Irrigated Areas Using Time-series AVHRR Pathfinder Datasets in the Krishna River Basin, India." *Photogrammetric Engineering and Remote Sensing* 73: 1029–1040.
- Thenkabail, P. S., V. Dheeravath, C. M. Biradar, O. R. P. Gangalakunta, P. Noojipady, C. Gurappa, M. Velpuri, M. Gumma, and Y. Li. 2009b. "Irrigated Area Maps and Statistics of India Using Remote Sensing and National Statistics." *Remote Sensing* 1: 50–67. doi:10.3390/rs1020050.
- Thenkabail, P. S., and Z. Wu. 2012. "An Automated Cropland Classification Algorithm (ACCA) for Tajikistan by Combining Landsat, MODIS, and Secondary Data." *Remote Sensing* 4: 2890. doi:10.3390/rs4102890.
- Thenkabail, P. S., M. A. Hanjra, V. Dheeravath, and M. K. Gumma. 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: 211–261. doi:10.3390/rs2010211.
- Vörösmarty, C. J., P. Green, J. Salisbury, and R. B. Lammers. 2000. "Global Water Resources: Vulnerability from Climate Change and Population Growth." *Science* 289: 284–288. doi:10.1126/science.289.5477.284.
- Vuolo, F., W.-T. Ng, and C. Atzberger. 2017. "Smoothing and Gap-filling of High Resolution Multi-spectral Time Series: Example of Landsat Data." *International Journal of Applied Earth Observation and Geoinformation* 57: 202–213. doi:10.1016/j.jag.2016.12.012.
- Waldner, F., G. S. Canto, and P. Defourny. 2015. "Automated annual cropland mapping using knowledge-based temporal features." *ISPRS J. Photogrammetry and Remote Sensing* 110: 1–13. doi:10.1016/j.isprsjprs.2015.09.013.
- Waldner, F., S. Fritz, A. Di Gregorio, D. Plotnikov, S. Bartalev, N. Kussul, P. Gong, P. S. Thenkabail, G. Hazeu, I. Klein, F. Low, J. Miettinen, V. Dadhwal, C. Lamarche, S. Bontemps, and P. Defourny. 2016. "A Unified Cropland Layer at 250 M for Global Agriculture Monitoring." *Data*. doi:10.3390/data1010003.
- WorldBank. 2015. "PovcalNet." Accessed 25 May 2015. <http://iresearch.worldbank.org/PovcalNet/povcalNet.html>
- WorldBank. 2018. "Consumer Price Indices Used in Global Poverty Measurement." *Global Poverty Monitoring Technical Note*. Accessed 03 February 2019. <http://povertydata.worldbank.org/poverty/region/SAS>
- Xiao, X., S. Boles, S. Frolking, C. Li, J. Y. Babu, W. Salas, and B. Moore. 2006. "Mapping Paddy Rice Agriculture in South and Southeast Asia Using Multi-temporal MODIS Images." *Remote Sensing of Environment* 100: 95–113. doi:10.1016/j.rse.2005.10.004.
- Xiong, J., P. Thenkabail, J. Tilton, M. Gumma, P. Teluguntla, A. Oliphant, R. Congalton, K. Yadav, and N. Gorelick. 2017a. "Nominal 30-m Cropland Extent Map of Continental Africa by Integrating Pixel-Based and Object-Based Algorithms Using Sentinel-2 and Landsat-8 Data on Google Earth Engine." *Remote Sensing* 9: 1065. doi:10.3390/rs9101065.
- Xiong, J., P. S. Thenkabail, M. K. Gumma, P. Teluguntla, J. Poehnelt, R. G. Congalton, K. Yadav, and D. Thau. 2017b. "Automated Cropland Mapping of Continental Africa Using Google Earth Engine Cloud Computing." *ISPRS Journal of Photogrammetry and Remote Sensing* 126: 225–244. doi:10.1016/j.isprsjprs.2017.01.019.
- Zhu, Z., S. Wang, and C. E. Woodcock. 2015. "Improvement and expansion of the Fmask algorithm: cloud, cloud shadow, and snow detection for Landsats 4-7, 8, and Sentinel 2 images." *Remote Sensing of Environment* 159: 269–277. doi:10.1016/j.rse.2014.12.014.