September 2017

Version

NASA Making Earth System Data Records for Use in Research Environments (MEaSUREs) Global Food Security-support Analysis Data (GFSAD) @ 30-m for North America: Cropland Extent Product (GFSAD30NACE)

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

The Global Food Security-support Analysis Data 30-m (GFSAD30) Cropland Extent-Product of North America (GFSAD30NACE) was produced by the following team members:

Mr. Richard Massey, PhD student at the Northern Arizona University, led the GFSAD30NACE mapping effort. His contributions include algorithm design, writing the code, computing and analysis, coordinating with the validation team, writing manuscripts, ATBDs, and user documentation.

Dr. Temuulen T. Sankey, Assistant Professor, Northern Arizona University, was the co-investigator working on the GFSAD30NACE product for North America. Her contributions include algorithm design, analysis, coordinating with the validation team and writing & editing the manuscript.

Ms. Kamini Yadav, PhD student at the University of New Hampshire was a lead member of the independent accuracy assessment team led by Dr. Russell G. Congalton.

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 GFSAD30NACE 30m cropland extent product for North America.

Dr. James C. Tilton, Computer Engineer with the Computational and Information Sciences and Technology Office (CISTO) of the Science and Exploration Directorate at the NASA Goddard Space Flight Center, developed recursive hierarchical segmentation (RHSeg) algorithm. He also helped Mr. Massey implement RHSeg on Northern Arizona University's computing cluster Monsoon.

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 conceptual framework of the GFSAD30 project and the GFSAD30NACE North America cropland extent product. He also provided constant intellectual guidance throughout the project, edited manuscripts, ATBD's, and user documentation.

II.   Historical Context and Background Information

Accurate spatial information on croplands is critical for global food security research, agricultural planning, and land-cover change studies (Foley et al. 2011; Thenkabail et al. 2010). Satellite image-based cropland maps provide spatially explicit, economic, and efficient methods and opportunities for cropland monitoring (Yu et al., 2013, Foley et al., 2011; Fritz et al., 2015; Wardlow and Egbert, 2008). North America provides much of the global crop production in the world. It is the largest producer of coarse grains and maize, and the third largest producer of cereals and wheat (Cerquiglini et al., 2016). A fundamental description of food production and food security as well as an indicator of the food supply system health are achieved by mapping and quantifying the spatial extent of croplands and can be used for economic and policy decision

making (Foley et al., 2011; Thenkabail et al., 2009). This algorithm theoretical basis document (ATBD) provides a detailed description of the GFSAD30 cropland extent product for all the 23 countries and the 9 independent territories in the North American continent (Table 1).

Table 1: Specifications for the global food security support-analysis data at 30m cropland extent product for North America (GFSAD30NACE).

Product name	Short name	Spatial resolution	Temporal resolution
GFSAD 30m Cropland extent product for North America	GFSAD30NACE	30m	Circa 2010

III. Rationale for Development of the Algorithms

The North American croplands have been mapped in many global land cover studies or cropland databases: croplands for 1992, 2001, 2006, and 2011 in the United States (US) National Land Cover Database (NLCD) (Homer et al., 2004; Homer et al., 2007; Homer et al., 2012), global irrigated and rainfed cropland maps and statistics (Thenkabail et al., 2009; Thenkabail et al., 2011), global cropland and pasture map (Ramankutty et al., 2008; Monfreda et al., 2008), land cover and land use (Goldewijk et al., 2011), global irrigated and rainfed areas (Portmann, et al., 2010), MODIS-derived global cropland extent (Pittman et al., 2010), 30m global cropland map (Yu et al., 2013), and MODIS global land cover product (Friedl et al., 2010). In addition, croplands have been mapped at both regional and country scales across the North American continent using remotely-sensed and ground-based data such as United States Department of Agriculture cropland data layer (CDL) (Johnson and Mueller, 2010; Boryan et al., 2011; Han et al., 2012; Boryan et al., 2014) and Agriculture and Agri-Food Canada annual crop inventory (Fisette et al., 2013; Fisette et al., 2014).

The current cropland maps and extents for North America, have significant shortcomings including: 1) coarse spatial resolution (250m pixels or larger), 2) low overall accuracies, 3) large disagreements in spatial distribution of croplands among current maps and with the country cropland area statistics, and 4) limited to country scales. We address these shortcomings in this study by providing a comprehensive and consistent cropland extent map GFSAD30NACE across the entire North American continent at 30m spatial resolution.

IV. Algorithm description

We leveraged the Google Earth Engine (Gorelick et al., 2017) (GEE)'s computing capacity and the Landsat data archive in this study to classify cropland extent at the North American continental scale. We used the Random Forest classification on GEE to develop a pixel-based classification of the North American continental cropland extent. We then fused the pixel-based classification with crop field boundaries across the North American continent identified using recursive hierarchical segmentation (RHSeg) (Tilton et al., 2012), an object-based classification

method on Northern Arizona University (NAU)'s high performance computing cluster Monsoon. This fusion of the two approaches resulted in a cropland extent map at 30 m spatial resolution for the North American continent for the nominal year 2010. We validated the final fusion-based cropland extent map using United States Department of Agriculture (USDA) cropland data layer (CDL) for the United States (US), Agriculture and Agri-food Canada (AAFC) annual crop inventory (ACI) in Canada, Servicio de Información Agroalimentaria y Pesquera (SIAP)'s digitized agricultural boundaries in Mexico, and high-resolution images from GEE application programming interface (API). Additionally, we validated the cropland extent map using USDA county crop statistics data, AAFC agricultural census data, and agricultural statistics from other countries in North America (Figure 1).
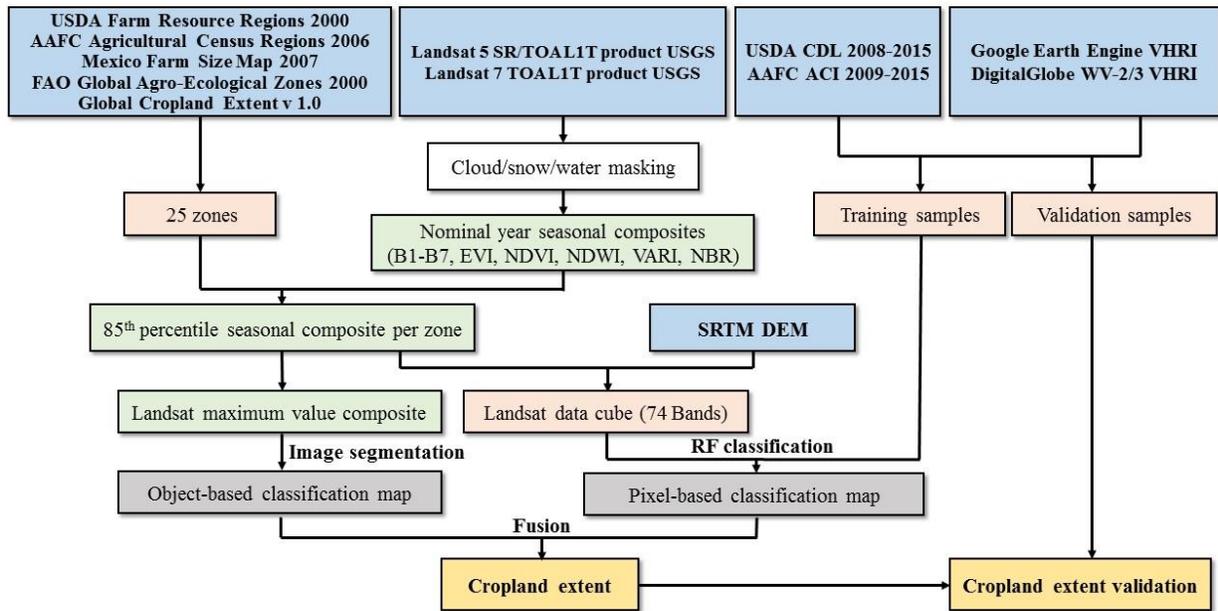
Figure 1: The data inputs and workflow for the object-based and pixel-based classification of the North American croplands in 30 m resolution and their fusion for cropland extent. The pixel-based Random Forest (RF) classification workflow was implemented on Google Earth Engine (GEE), while the Recursive Hierarchical segmentation (RHSeg) object-based classification and fusion of the object-based and pixel-based classification outputs were implemented on Northern Arizona University's high-performance computing cluster Monsoon.
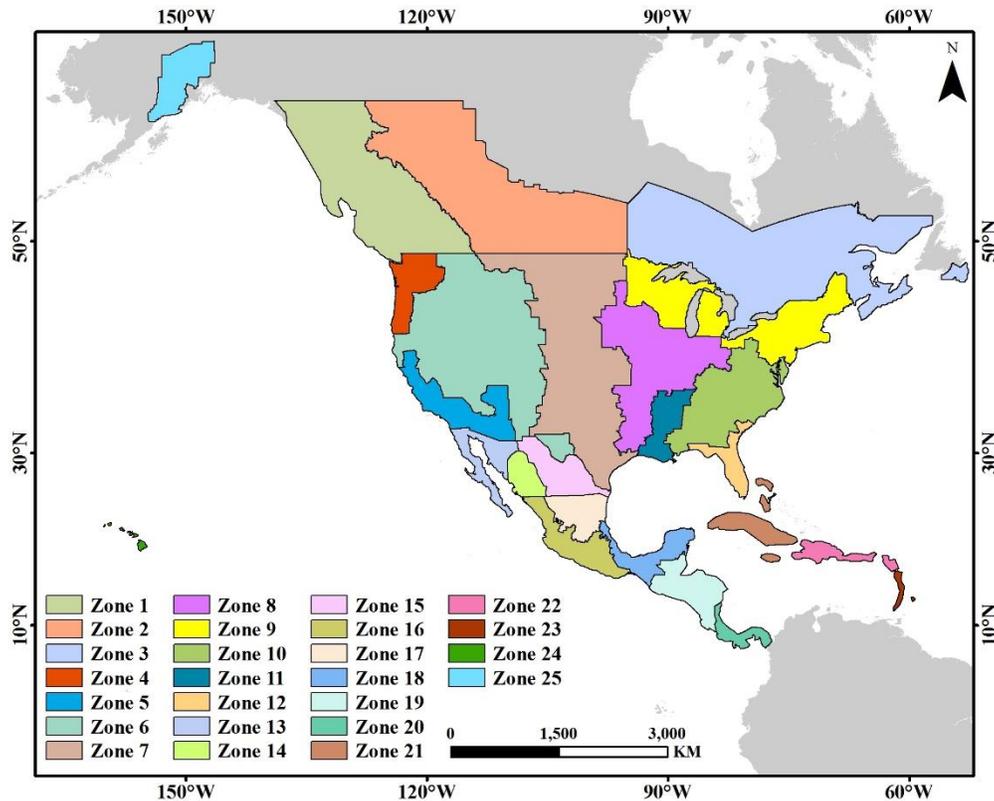
Figure 2: The study area of North America. The entire North American continent was divided into 25 zones of similar farming practices using the US Department of Agriculture (USDA) Farm Resource Regions 2000 (FRR), Canadian Agriculture Regions with 2011 agricultural census, Canada Vegetation Regions 1998, Mexico farm sizes from Instituto Nacional de Estadística y Geografía (INEGI) 2007, and the Food and Agriculture Organization (FAO) global Agro-Ecological Zones 2000 (AEZ).

A. Input data

1. Study area

Our study area included the entire North American continent, which is divided into five major regions by political and physical boundaries: 1) Canada, 2) the US, 3) Mexico, 4) Central America, and 5) the Caribbean islands. The five regions were further divided into 25 zones (Figure 2) to identify areas of similar farming practices within similar global agro-ecological zones (AEZs) defined by the FAO based on the number of growing degree days, soil, and terrain data in 10 km spatial resolution (Fischer et al., 2000). The division of regions into zones was necessary to reduce variability in: 1) satellite image sensors used, 2) availability of surface reflectance data, 3) satellite image-derived input variables, and 4) length and dates of periods for input variables.

Canada included three zones based on the 2011 Canadian agricultural census regions contained within major Canadian Vegetation Regions (Wulder et al., 2008). The conterminous US was divided into 9 zones based on the USDA Farm Resource Regions (FRR). The FRR represent areas with similar types of farms intersected with areas of similar physiographic, soil, and climatic variables (Economic Research Service, 2000). The Alaska zone was limited to the counties in the Cook Inlet lowlands. The Hawaii zone included all of the Hawaiian Islands in the archipelago. Similarly, Mexico was divided into six zones based on the average farm sizes of its states (INEGI, 2007) and the availability of Landsat surface reflectance data in the GEE data archive. The Central American region included two zones, while the Caribbean islands were divided into three zones of major island groups. Additionally, we used the Global Cropland Extent (GCE) v1.0 map (Teluguntla et al., 2015) to determine the initial overall extent of croplands in the North American continent to exclude areas without any cropland such as northern Canada.

For both the pixel-based RF and the object-based classifications, each of the zones in Canada, the US, and Mexico were further sub-divided into smaller sub-zones. The division of the 25 zones into sub-zones was also necessary to: 1) streamline the classification process in Google Earth Engine programming interface (GEE API) by limiting the amount of training samples processed in one instance, 2) further reduce training sample spatial variability, 3) for rapid visualization of the classification output to iteratively improve training sample sets in each zone, and 4) to minimize variability in mean field sizes for the object-based classification. The zones and the sub-zones for the North American continent were imported in GEE API by converting sub-zone shapefiles to Google fusion tables.

2. Processing platforms

We used Google Earth Engine (GEE) for the pixel-based cropland classification. The Landsat archive available on GEE is already pre-processed for atmospheric and topographic effects, which saved us much effort in data download and pre-processing times. We used JavaScript in the GEE code editor, which is an application programming interface (API). The zone boundaries were imported into GEE via Google fusion tables. The training and validation samples were generated using the Map tab of the GEE code editor.

We used Northern Arizona University's (NAU) high performance computing cluster, Monsoon (NAU Monsoon) for: 1) compiling the pixel-based classification results from GEE, 2) performing an object-based classification of the Landsat composites using the Recursive Hierarchical segmentation (RHSeg) software (Tilton et al., 2016; Tilton, 2012), and 3) fusion of the pixel-based classification output from GEE and the object-based classification output. We used IDL 8.5 (Exelis Visual Information Solutions, Boulder, Colorado) as the scripting language for parallel processing the Landsat composites via the RHSeg software. Additional post-processing of the crop field boundaries was performed also on NAU Monsoon using IDL 8.5. A brief overview of this process is shown in Figure 2.
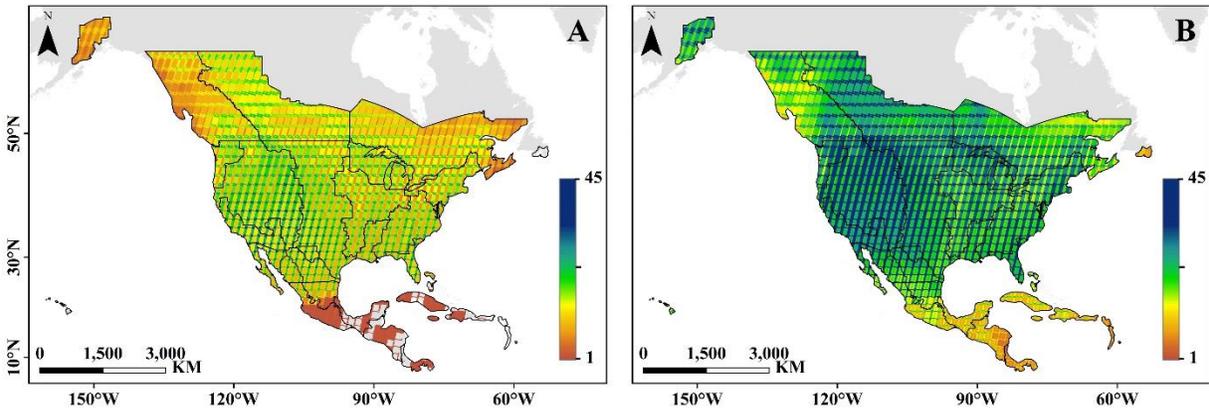
Figure 3: The average yearly availability of Landsat 5 TM surface reflectance scenes (panel A) and the combination of Landsat 5 TM and Landsat 7 ETM+ top-of-atmosphere scenes (panel B) for the study zones in North America for the years 2008-2012 on Google Earth Engine.

3. Satellite Images

The Landsat 5 Thematic Mapper (TM) surface reflectance (SR) image collection available in GEE image archive was used for this study. Specifically, the image collection LANDSAT/ LT5_SR for 2008 – 2012 was used for the nominal year 2010. The Landsat SR data available in GEE are already processed using the Landsat ecosystem disturbance adaptive processing system (LEDAPS) algorithm (Masek et al., 2012) and includes a cloud mask ("cfmask") band, which we used to mask the cloud pixels. In the zones where SR data wasn't available, for example, zones 17-25 in southern Mexico, the Central American countries, the Caribbean, Hawaii, and Alaska (Figure 3), we used a combination of top-of-atmosphere (TOA) Landsat 5 TM and Landsat 7 Enhanced Thematic Mapper Plus (ETM+) data which were available as GEE data collections "LANDSAT/LT5_L1T_TOA" and "LANDSAT/LE7_L1T_TOA", respectively. While GEE uses Spherical Normal Mercator projection for display in its JavaScript API maps tab, it automatically re-projects satellite remote sensing data to a common projection inside its servers during processing. All of the Landsat data composite prepared via GEE and downloaded to NAU Monsoon for object-based classification were projected in Geographic projection with WGS-84 datum and 30m pixels. In addition, elevation and slope data from the Shuttle Radar Topography Mission (SRTM) 3.0 digital elevation model (DEM) at 1 arc-second were used for topographic information.

We used the multi-year Landsat data for the entire continent to: 1) maximize data availability in all regions, 2) maximize the spatial extent of croplands mapped, and 3) reduce the effects of cloud cover. The nominal year was further divided into multiple intervals or periods (Figure 4) depending on the seasonal variability in the region and availability of the cloud free Landsat data (Table 2).

9

Table 2: Days of the year and number of periods in the nominal year 2010 for study area zones in North America.

Zones	Number of periods	Day of year
1, 2, 3, 25	4	1-90, 91-180, 181-270, and 271-365
4, 5, 6, 7, 8, 10, 11, 12, 13, 14, and 15	6	0-60, 61-120, 121-180, 181-240, 240-300, and 301-365
16, 17, and 18	4	1-90, 91-180, 180-270, and 270-365
19 and 20	3	1-60, 61-180, and 181-365
21	3	60-120, 121-180, and 181-270
22, 23, and 24	3	1-60, 61-180, and 181-365

Cloud cover and cloud shadows present major difficulties in cropland classification in south Mexico, Central America, and the Caribbean islands. We eliminated clouds and cloud shadows for each TOA Landsat scene in these regions by using a combination of cloud likelihood function, *simpleCloudScore,* on GEE, and temporal dark outlier mask (TDOM) method (Housman et al., 2015). The *simpleCloudScore* function uses a combination of brightness temperature and normalized difference snow index (NDSI) values to identify likelihood of a pixel being a cloud pixel, while the TDOM method identifies pixels that are dark in bands 4, 5, and 6 in the given Landsat scene by finding statistical outliers with respect to the average of bands 4, 5, and 6 at the same location on different dates. While both methods can be used to derive cloud and cloud shadow masks for each Landsat scene, the TDOM method requires > 4 scenes at a Landsat footprint to be effective in identifying cloud shadows. We, therefore, used the combination of both methods.
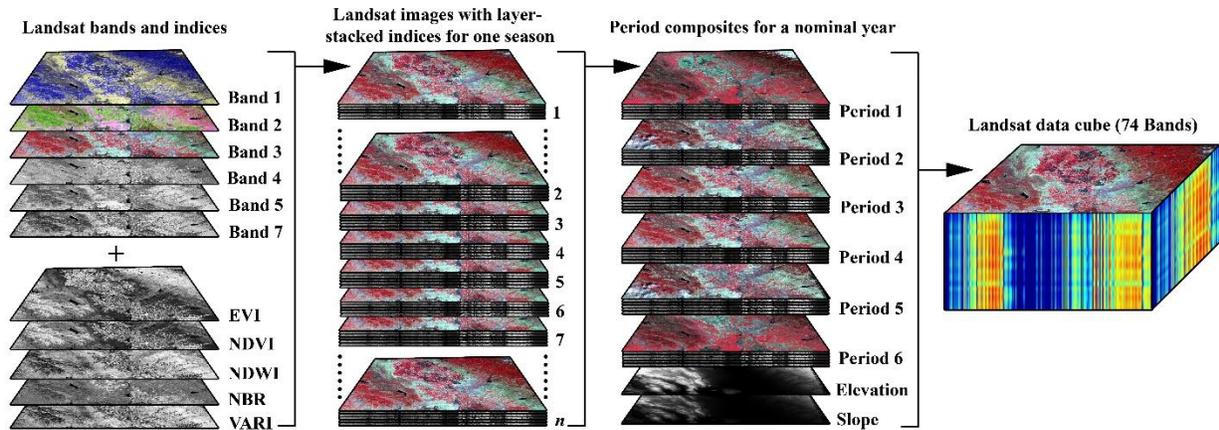
Figure 4: An example of input bands used to create a Landsat data cube for the Random Forest (RF) classification approach in the nominal year 2010. The periods 1-6 include days of year ranging: 1-60, 61-120, 121-180, 181-240, 240-300, and 301-365, respectively.

4. Input variables

We developed three major types of input variables for the Random Forest (RF) classification: 1) spectral, 2) phenological, and 3) topographic variables. The spectral variables included Landsat bands 1, 2, 3, 4, 5, and 7 for each period. A single mosaicked image was produced for each period by reducing the image collection to the 85th percentile enhanced vegetation index (EVI) value for each pixel in all bands and indices across all available image dates for the period. In areas with substantial parts of the year under snow cover, we used the 85th percentile NDVI values. The 85th percentile values eliminated cloud and snow pixels that could not be eliminated in the surface reflectance by the LEDAPS algorithm and at the same time maximized the number of cropland pixels. Additionally, in Mexico and Central America regions, the 85th percentile EVI values effectively removed clouds and minimized Landsat 7 scan-line correction (SLC) artifacts in the Landsat composites. This resulted in a total of 36 spectral band input variables: Landsat bands 1, 2, 3, 4, 5, and 7 for six periods for the nominal year. In zones with three and four periods, the total number of spectral variables were 18 and 24, respectively. Periods with non-contiguous data due to clouds were removed from the input data.

The phenological variables included: 1) enhanced vegetation index (EVI) (Heute et al., 2002), 2) normalized difference vegetation index (NDVI) (Rouse et al., 1972), 3) normalized difference water index (NDWI) (McFeeters, 1996), 4) visible atmospheric resistant index (VARI) (Gitelson et al., 2002), and 5) normalized burn ratio (NBR) (Key and Benson, 1999). We calculated NDVI and NDWI for all periods. NDVI time-series provides critical crop phenology information. While the nominal year in our study included six periods or less, the NDVI time-series provided sufficient sensitivity to crop phenology to differentiate it from non-crop pixels. NDWI is a good indicator of vegetation water content and has low sensitivity to atmospheric scattering effects (Gao et al., 1996). NDWI is calculated similarly to NDVI, but using Landsat image bands 4 and 5. NDWI of crop fields show higher negative anomalies during sowing than other vegetation types. Furthermore, the standard deviation of NDWI is higher for crop pixels than non-crop vegetation pixels.

VARI is based entirely on the visible range of the spectrum and has low sensitivity to atmospheric effects. It was developed to estimate vegetation fraction (Schneider et al., 2008). Unlike NDVI, VARI shows a linear response to vegetation fraction for its entire range (Gitelson et al., 2002). EVI and VARI were particularly useful in differentiating crop pixels from non-crop vegetated pixels in areas with high rainfall. NDVI and NDWI both saturated as vegetation abundance increased in these areas. Additionally, we used NBR to highlight sensitivity to vegetation water content in highly wet and highly dry regions (Key and Benson, 1999). NDVI, EVI, NDWI, VARI and NBR were computed using the following equations:

$$NDVI = \frac{B4 - B3}{B4 + B3} \qquad (1)$$

$$EVI = \frac{B4 - B3}{B4 + 6 \times B3 - 7.5 \times B1 + 1} \qquad (2)$$

$$NDWI = \frac{B4-B5}{B4+B5} \qquad (3)$$

$$VARI = \frac{B2-B3}{B2+B3-B1} \qquad (4)$$

$$NBR = \frac{B4-B7}{B4+B7} \qquad (5)$$

where B1, B2, B3, B4, B5, and B7 represent Landsat blue, green, red, near infrared, and shortwave infrared 1 & 2 bands, respectively. We used all the phenological input variables to maximize cropland and non-cropland classification accuracy.

The topographic variables included slope and elevation derived from the United States Geological Survey (USGS) SRTM (Farr et al., 2007) digital elevation model (DEM) at 1 arc-second (30m). As most crops are cultivated on flat lowlands, slope is a prominent variable in large scale cropland mapping. Similarly, the range of elevation variation was narrow for croplands and provides a critical input.

We also calculated two period maximum value composite (MVC) based on the 85th percentile value of NDVI with Landsat bands 1-7 for all regions in North America with Julian date ranges 60-180 and 181-270 as input variables for the object-based classification in RHSeg. The two-period input introduced minor variability in spectral outputs for adjacent crop fields, which allowed successful classification as two different objects rather than one large contiguous object. The MVC was also used to inform the training and validation sample sets to include all spectral variability across each sub-zone for the nominal year.

5. Reference data

Reference data were obtained for the entire North American continent for training and testing. An independent validation dataset was generated for accuracy assessment of the classification output. We used available reference cropland layers in Canada, the US, and Mexico to derive reference data for training and testing. In other countries, it was necessary to collect or generate the appropriate reference data as none had been previously collected.

We used USDA CDL to derive the reference data in the US. The USDA CDL was produced in 56 m spatial resolution for 2008 and 2009 in some states, but in 30 m resolution for all US states after 2010 (Johnson and Mueller, 2010; Boryan et al., 2011; Han et al., 2012; Boryan et al., 2014). The 2008-2014 CDL maps were re-projected to geographic latitude/longitude in WGS 1984 datum and 0.00027 degree (approximately 30 m) spatial resolution using nearest neighbor resampling. A cropland extent reference map for the US was then produced for the nominal year 2010 by binary overlay of cropland extents for years 2008-2014. In the binary overlay for the nominal year, any pixel that was labeled once or more times as cropland or as fallow in the USDA CDL was considered cropland. All other pixels were considered non-cropland. We considered all field crops, tree crops, alfalfa, other hays, and plantation crops as cropland. The cropland extent derived from the statewide agricultural land-

use baseline 2015 data (Melrose et al., 2015) using similar crop categories was used as the reference data for Hawaii.

The AAFC ACI was used to derive reference data in Canada. The AAFC ACI was produced in 56m spatial resolution in 2009 and 2010 for some provinces, but in 30m resolution for all subsequent years (Fisette et al., 2013; Fisette et al., 2014). Additionally, the ACI was produced for all provinces in Canada from 2012 onwards. The 2009-2015 crop inventory layers were re-projected to geographic latitude/longitude in WGS 1984 datum and 0.00027 degree (30 m) pixel size using nearest neighbor resampling. The cropland extent maps in Canada for 2010 was then produced similarly to the US dataset by binary overlay of the crop inventory layers.

In Mexico, we used the SIAP agricultural frontier boundaries of agricultural land use to derive reference data. This reference data includes cropland extent classified using SPOT 5 satellite imagery from 2010-2011 (SIAP, 2017). No other reference data were available to us for training our classification for the other regions in North America. In the countries where no previous reference data were available, very high-resolution images (VHRI) available on GEE and from DigitalGlobe WorldView-2 data was used to generate training datasets.
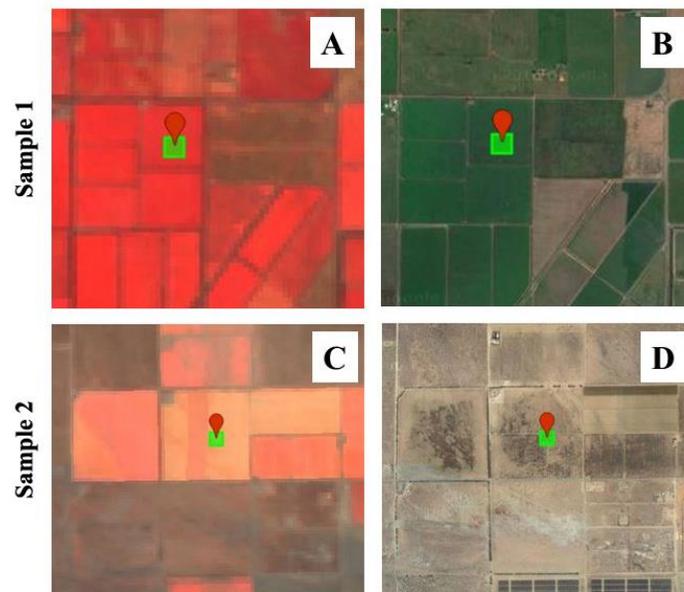
Figure 5: Sample selection using Landsat generated maximum-value composites (panels A and C) and Google Earth Engine high resolution imagery (panels B and D) at two different locations. Panels A and C show visual differences in color infrared images of the sample locations.

The training and testing data generated for classification was kept separate and independent from the validation samples generated for accuracy assessment. For training, approximately 800 polygons, each 90m x 90m in dimension, were randomly generated as training sample units for cropland and for non-cropland cover types within each sub-zone. Pixels inside each 90m x 90m polygon were averaged to give one value per sample in each band. The

13

training samples from each zone were visually assessed and compared with Landsat MVC to ensure that the samples included all spectral variations of the cover type. For example, difference in crop types may introduce variations in the color infrared images (Figure 5). We included approximately all such visual variations in the color infrared images of crop class in each zone. The total number of training and validation samples in zones (Figure 2) and sub-zones (Figure 6) are listed in Table 3.

Table 3: Training and validation sample summary for sub-zones in North America.

Country / Region	Zone	Sub-zone	Sub-zone name	Training Samples		Validation Samples	
				Crop	Non-crop	Crop	Non-crop
Canada	1	1	Yukon	334	487	1	249
	2	1	NWForest	530	600	73	177
	2	2	CanPrairies	873	1,081		
	3	1	CenForest	184	224	10	240
	3	2	MidEasternForest	389	455		
	3	3	NEasternForest	46	64		
	3	4	NovaScotia	744	983		
	3	5	NewFoundland	72	112		
United States	4	1	NWRim	694	910	27	223
	5	1	SWRim	1,016	1,255	29	221
	6	1	WBasinRange	555	679	12	238
	6	2	WRim	350	364		
	6	3	NBasinRange	460	452		
	6	4	SBasinRange	468	475		
	6	5	SRim	393	402		
	7	1	NGreatPlains	936	1,119	77	173
	7	2	SGreatPlains	781	719		
	7	3	NPrairies	512	637		
	7	4	SPrairies	724	883		
	7	5	SERim	672	648		
	8	1	Heartland	689	551	103	147
	8	2	MidUplands	393	483		
	8	3	MidSeaBoard	210	304		
	9	1	WCrescent	567	690	31	219
	9	2	MidCrescent	628	848		
	9	3	ECrescent	1,261	1,226		
	10	1	EUplands	1,013	1,007	52	198
	10	2	ESeaBoard	1,122	1,173		
	11	1	Mississippi	571	628	66	184
	12	1	ERim	855	1,089	31	219

Country							
Mexico	13	1	BajaSonora	243	275	12	237
	13	2	Baja	130	249		
	14	1	Sonora	405	703	67	181
	15	1	Chihuahua	824	1,072	20	227
	16	1	SMexBasinRange	1,024	1,249	58	183
	16	2	Mexico	1,147	956		
	16	3	Guerrero	1,011	895		
	17	1	SMadre	653	845	25	224
	18	1	Veracruz	789	716	50	179
	18	2	Yucatan	687	613		
Belize, Guatemala, El Salvador, Honduras	19	1	CenAM1	2,320	2,436	69	181
Costa Rica, Panama	20	1	CenAM2	1,329	1,445	52	194
Bahamas	21	1	Freeport	45	56	77	169
	21	2	Bahamas	157	328		
Cuba	21	3	Cuba	716	732		
Jamaica	21	4	Jamaica	436	526		
Dominican Republic, Haiti, Puerto Rico	22	1	DRHaiti	1,203	1,174	74	174
British Virgin Is., US Virgin Is., Anguilla, St. Kitts & Nevis, Antigua & Barbuda, Montserrat, Saba, Saint Barthelemy, Saint Martin, Sint Eustatius, and Sint Marteen	22	2	LCarIS	461	419		
Guadeloupe, Dominica, Martinique, St. Lucia, Grenada, St. Vincent and the Grenadines	23	1	CarIS	396	458	43	203
Barbados	23	2	Barbados	89	72		
Hawaii	24	1	Hawaii	604	616	40	210
Alaska	25	1	Anchorage	165	204	4	246
Total				**32,876**	**36,587**	**1,103**	**5,096**

Figure 6: Sub-zones used for the pixel-based and object-based classification in the North American continent.
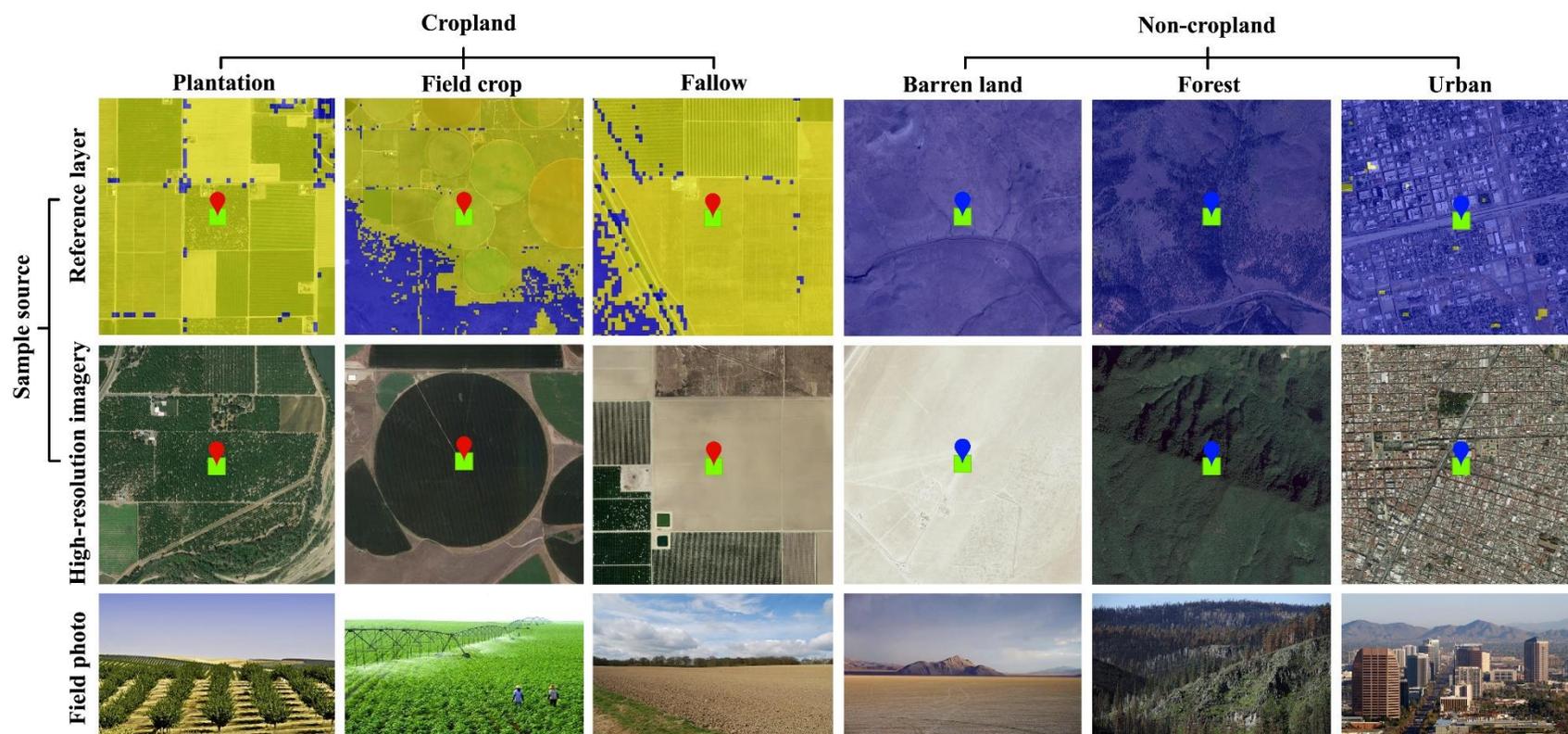
Figure 7: Examples of land cover categories used to derive training and validation samples for pixel-based random forest classification on Google Earth Engine. The samples for crop (red markers) were selected in the center of fields within a 90m × 90m square. The samples were generated randomly and assigned labels after interpreting high resolution satellite imagery.

The training samples were also assessed by comparing with the available layers, the USDA CDL, AAFC ACI, and SIAP agricultural frontier boundaries in the US, Canada, and Mexico, respectively (Figure 7). The total numbers of training samples for the entire North American continent were 32,876 and 36,587 for the crop and non-crop classes, respectively. In addition, random samples were also generated by digitizing crop field boundaries for all zones in the North American continent to test the RHSeg-derived object-based classification. These samples were selected in crop locations not included in the training data set by interpreting the VHRI in GEE and the reference data available for the US, Canada, and Mexico.

An independent validation sample dataset was also collected for statistical accuracy assessment. We selected 250 validation samples in each zone. These samples were randomly selected from locations not used by the mapping team for training or testing. We used existing reference data in the US, Canada, and Mexico and visual interpretation of VHRI in GEE to extract the validation samples from the center of crop fields. For all the other zones in North America, the validation samples were collected by visual interpretation of VHRI in GEE by two interpreters. To minimize the error in validation data collection, only those samples where both interpreters agreed with the label were used for accuracy assessment.

B. Theoretical description

1. Cropland definition

We classified cultivated land with plants harvested for food, feed, and fiber (e.g., wheat, rice, corn, soybeans, cotton) and continuous plantations (e.g., coffee, tea, rubber, cocoa, and oil palms) as croplands (Table 4). We also included fallow lands that are uncultivated during a period or a year but are equipped for cultivation, as croplands. Non-croplands include all other land cover types.

Table 4: Dominant crop types and classes in the North American continent classified as crop.

Corn	Mustard	Orchards
Cotton	Canola	Potatoes
Rice	Sunflower	Tomatoes
Sorghum	Soybeans	Vegetables
Barley	Oil palm	Beans
Durum Wheat	Safflower	Lentils
Spring Wheat	Peanuts	Tea
Winter Wheat	Sugarcane	Coffee
Rye	Grapes	Cocoa
Oats	Berries	Rubber
Millet	Fruits	Alfalfa and other hay

2. Algorithms

We used the Random Forest classification on GEE to develop a pixel-based classification of the North American continental cropland extent. We then fused the pixel-based classification with crop field boundaries across the North American continent identified using recursive hierarchical segmentation (RHSeg) (Tilton et al., 2012), an object-based classification method on Northern Arizona University (NAU)'s high performance computing cluster Monsoon. This fusion of the two approaches resulted in a cropland extent map at 30 m spatial resolution for the North American continent for the nominal year 2010.

C. Practical description

1. Random forest classification

Random Forest (RF) classification is a machine learning method, in which the RF classifiers construct multiple de-correlated random decision trees that are bootstrapped and aggregated to classify a dataset by using the mode of predictions from all decision trees (Breiman, 2001). The RF classifiers are generally immune to data noise and overfitting, and are extremely useful in classifying remote sensing data. Furthermore, RF classifiers typically achieve higher accuracies in comparison with other approaches such as maximum likelihood, single decision trees, and single layer neural networks (Lawrence et al., 2006; Na et al., 2010). RF classifiers also provide a quantitative measurement of each variable's contribution to the classification output, which is useful in evaluating the importance of each variable. RF classifiers provide an internal accuracy assessment by using an 'out-of-bag' (OOB) technique, in which about a third of the data is kept aside as validation dataset to assess accuracy of the classification. This technique can be used to cross-validate the RF classifier using independent datasets.

RF classifiers available in GEE use six input parameters: 1) number of classification trees, 2) number of variables used in each classification tree, 3) minimum leaf population, 4) bagged fraction of the input variables per decision tree, 5) out-of-bag mode, and 6) random seed variable for decision tree construction. When the number of trees increase, the overall accuracy of classification increases without overfitting (Breiman, 2001). While training sample imbalance can affect the RF classification output by over-fitting the majority class (Breiman, 2001; Chen, 2004), various methods such as down-sampling the majority class or duplicating the minority samples can provide immunity against over-fitting (Sun et al., 2007). Further, the choice of optimized parameter values using the OOB outputs can eliminate overfitting in the RF classifier.

We tested the number of variables per split from 1-100 in three different zones in the US by incrementally increasing the number by 5 using the 'tuneRF' function in the 'randomForest' package in R (Liaw, 2015). These three zones had different sample sizes: 1) 350 crop and 364 non-crop, 2) 628 crop and 848 non-crop, and 3) 405 crop and 703 non-crop samples. We found that the internal classification accuracy using OOB mode of the RF classifier did not change by more than 2% across the three zones. We, therefore, kept the number of variables per split to 5 as the performance at this value always appeared to be above average, while the computation times were the lowest. We set the number of trees to about a fourth of the maximum number of samples in the crop or non-crop class rounded to the next hundred.

The RF model accuracy was computed by using the OOB mode using the R 'randomForest' package to check for over-fitting. The analysis of the error rates from the OOB outputs yield consistent accuracies with low variations across independent datasets and hence validate the absence of over-fitting. Furthermore, in the zones where the training samples were not balanced, the samples in the minority class were carefully selected by visual assessment of all variations at training sample locations using MVCs and VHRI in GEE.

The selections of RF input variables were evaluated in comparison with the MVC to understand their contributions and changes in runtime and to choose the final set of input variables. It was found that adding a combination of EVI, NDVI, and NDWI bands achieved >90% RF model internal accuracy contributing to an increase in accuracy of 18% in semi-arid areas. On the other hand, adding all the spectral bands from all available periods resulted in an accuracy increase of 8%. Finally, adding all the phenological and spectral bands resulted in average internal accuracy of the RF model increase by 21%, while runtime was significantly shorter. In more humid areas, however, the increase in average internal accuracy was 20%, when we added VARI and NBR bands. Hence, EVI, NDVI, and NDWI bands were used along with the spectral bands, while VARI and NBR bands were further added in humid areas.

2. Recursive hierarchical segmentation

We used the Recursive Hierarchical Segmentation (RHSeg) software (Tilton et al., 2016; Tilton, 2012) version 1.64 to identify individual crop fields and further refine the pixel-based classification output from the RF approach (Figure 8). RHSeg is an image segmentation algorithm, which uses hierarchical step-wise optimization (HSWO) region growing approach (Beaulieu and Goldberg, 1989) by merging spectrally similar and adjacent regions together based on image texture (Tilton, 2016). While the HSWO approach can produce exhaustive segmentation hierarchies via region growing from initialization until the entire image tile becomes one region, Hierarchical Segmentation (HSeg) selects the most appropriate number of iterations by monitoring the merging thresholds between objects. These thresholds determine merging of multiple joint or disjoint segments into a single segment based on the dissimilarity values of image pixels. RHSeg computes an overall merge threshold for each segmentation hierarchy. The HSeg algorithm allows for merging of both spatially adjacent and non-adjacent regions providing flexibility and user-defined parameters in the final output, but leads to heavy computational demands. RHSeg, however, is a computationally efficient version of the Hierarchical Segmentation (HSeg) algorithm and recursively sub-divides the image tile into smaller sections to limit the number of regions being processed at a time (Plaza and Tilton, 2005). RHSeg can be implemented using multiple parallel processes that can process the sub-divided tiles in parallel and further improve the computational efficiency.
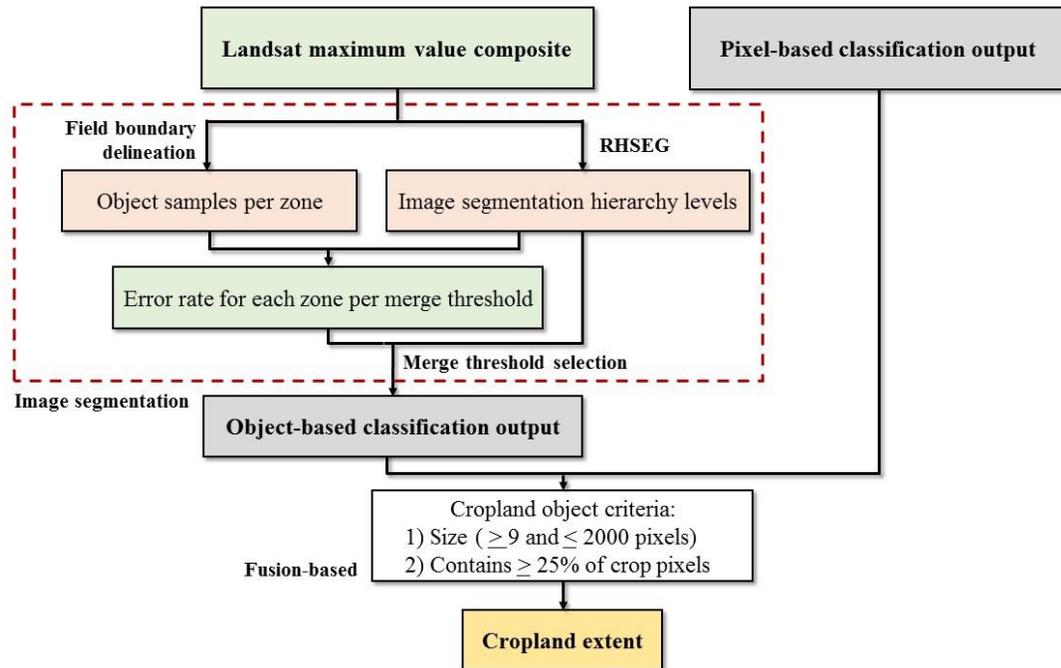
Figure 8: Object-based image classification workflow via the RHSeg software. The Landsat bands 1-7 median value composite from two are used as input variables. Sample field boundaries were delineated by digitizing high resolution imagery in Google Earth Engine (GEE) and were used to identify best hierarchy level for RHSeg, based on minimum error rates of comparison between objects corresponding to the sample field boundaries. The object-based classification output and RF pixel-based classification output are then fused together by using two criteria related to the object size and abundance of cropland pixels within each object.

We divided the MVC input variables for the North American continent into 1,898 tiles, 1.2 degree × 1.2 degree (110.5 km × 110.5 km) in dimension, which includes a buffer of 0.1 degree (11.05 km) for side overlap with adjacent tiles (Figure 9). These tiles were processed in parallel on NAU Monsoon computing cluster using custom scripts to implement RHSeg software. Any tile in which the maximum size of a cropland object was less than 9 Landsat pixels (90 x 90m) was discarded. We identified 25-40 sample objects (Figure 10) by manually delineating field boundaries in each zone, with up to 2-3 sample objects in the same tile, for selection of results in segmentation hierarchy. The selected sample objects were divided across each zone in well distributed tiles.

The major parameters for RHSeg program include: 1) initial region classes, the number of region classes at which segmentation hierarchy output was initiated, 2) spectral clustering weight, the weight of spectral clustering versus region growing, and 3) the dissimilarity criterion used to classify objects (Tilton, 2012). In addition to the above parameters, RHSEG hierarchy was output at selected merge thresholds.

21
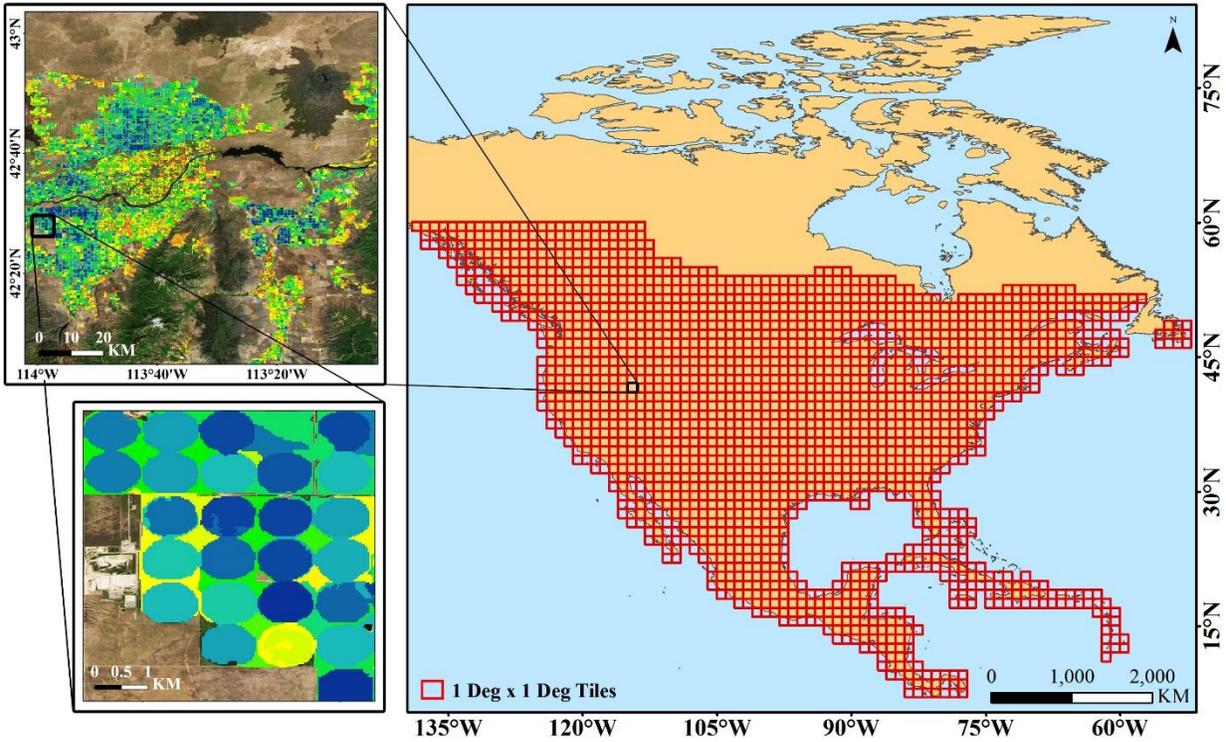
Figure 9: The North American continent was divided into 1,898 tiles, each 1.2 degree × 1.2 degree (132.6 km × 132.6 km) in dimension, which includes a 0.1 degree (11.05 km) side overlap, for the object-based classification of croplands using the Recursive Hierarchical Segmentation (RHSeg) software.

The RHSeg program begins with many small regions and recursively merges them together in each hierarchy to form meaningful object-based classifications (Figure 11). Every successive hierarchy, therefore, has fewer objects compared to the previous segmentation hierarchy. We used a value of 0.5 for the spectral clustering weight to perform spectral clustering of disjoint objects into region classes. We constrained the segmentation hierarchy using the initial region classes to start at 220,000 region classes. RHSeg program was terminated at 10,000 region classes per tile in the last segmentation hierarchy. As a result, our total number regions per tile could range from 10,000 to 220,000. As the spectral clustering was also allowed at 0.5, the number of objects was larger than the number of regions. The limits of 10,000-220,000 region classes in a tile were identified by assuming one object per class and identifying the maximum number of the largest and the smallest cropland objects as 2000 and 9 pixels respectively. The dissimilarity criterion used to identify different objects in the object-based classification was based on minimizing the increase in the mean squared error between the region mean image and the original image data (Tilton, 2016) and was indicated by the value of the dissimilarity criterion parameter as 6. RHSeg generated up to 40 hierarchy levels of image segmentations per tile.

22

3. Pixel-based and object-based classification fusion

We identified the optimum segmentation hierarchy for merging RHSeg output in each sub-zone with the pixel-based random forest classification. The number of hierarchy levels and the merge thresholds varied across tiles in a sub-zone. We calculated a common optimum segmentation hierarchy for a sub-zone by analyzing error rates across all the segmentation hierarchies in all the tiles and identifying the error rate with the lowest value and least variations with merge thresholds. We calculated error rates for each segmentation hierarchy using the number of committed and omitted object pixels in the sample objects. The RHSEG output objects with maximum overlap with the sample objects were selected to calculate the error rate in each segmentation hierarchy using the RHSeg 1.64 utility program *hsegrefcomp*. The *hsegrefcomp* program produced summary tables that include merge thresholds and their error rates for all the samples in the sub-zone. The error rate value was calculated for each merge threshold increment by binning the error rate values for all the samples within the increment range and calculating the median error rate for each bin. The merge threshold increment was calculated by dividing the difference in the maximum and minimum merge thresholds with the number of hierarchy levels.
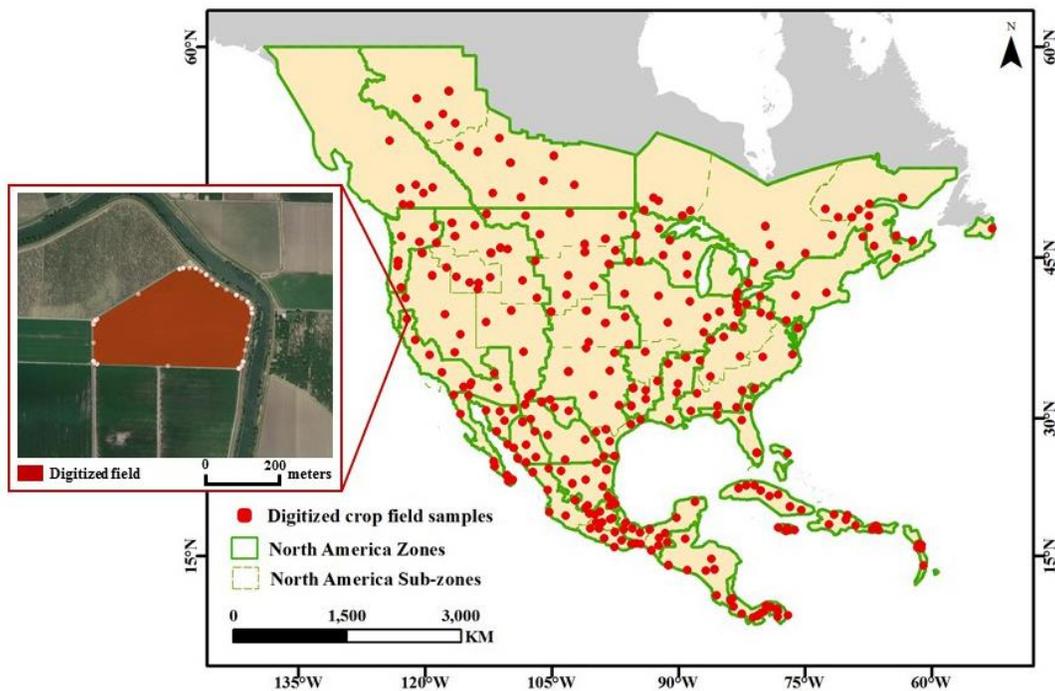
Figure 10: A total of 25-40 region objects were identified per sub-zone within each of the 25 zones (total n = 1,577) as digitized field boundaries for training the segmentation hierarchy in RHSeg.
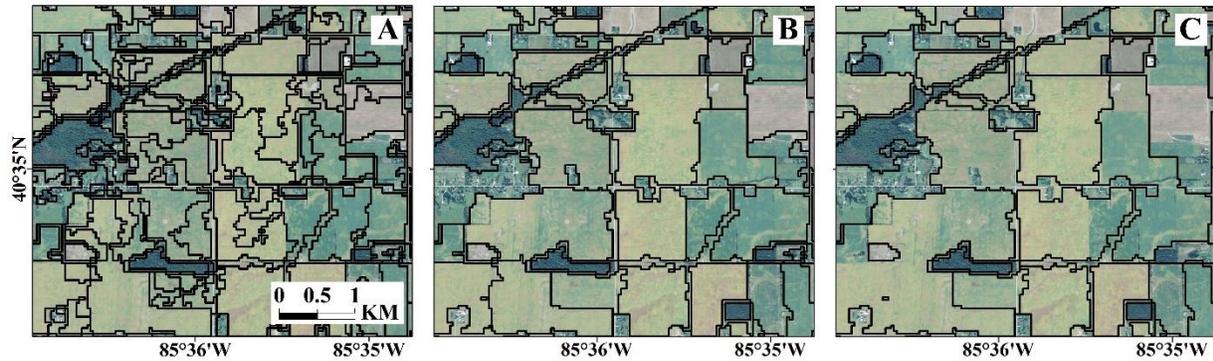
Figure 11: RHSeg-derived segment boundaries for the lowest segmentation hierarchy at level 1 (panel A), a medium segmentation hierarchy at level 21 (panel B), and the highest segmentation hierarchy at level 35 (panel C) for zone 8 in the Heartland region of the US. We selected level 21 as the optimum segmentation hierarchy in this zone.
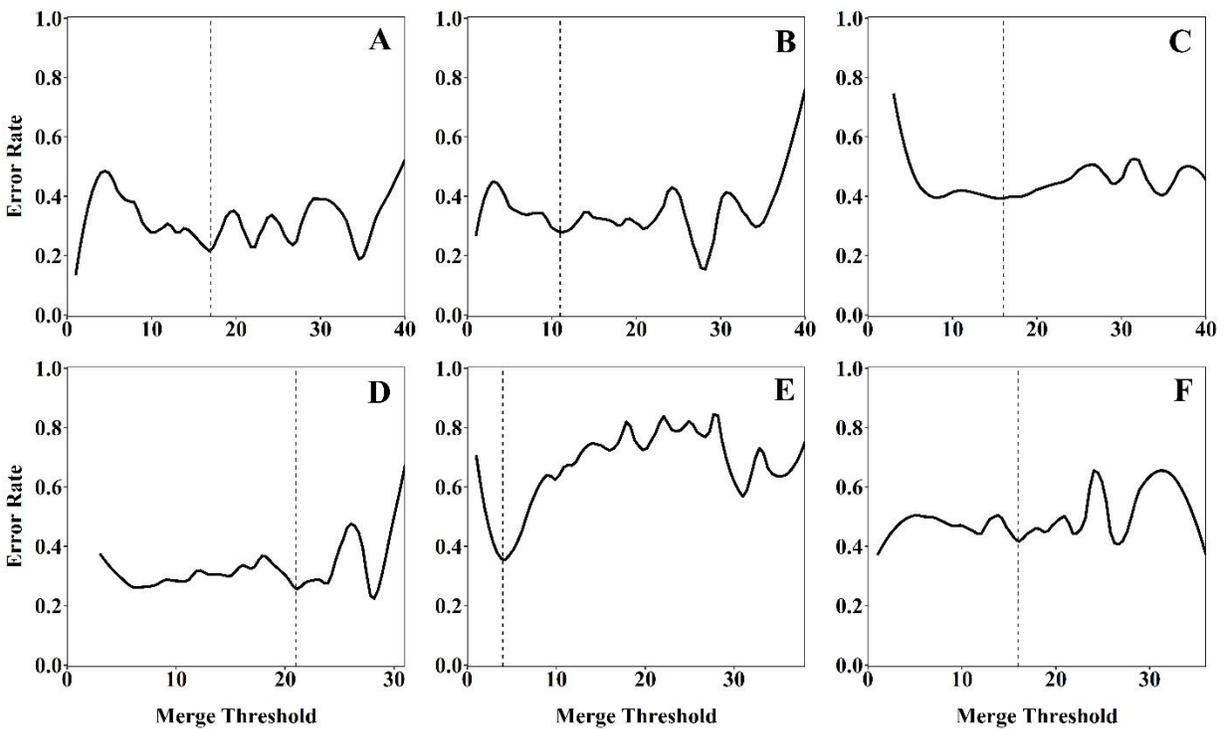
Figure 12: Error rates versus merge thresholds in different zones and sub-zones in North America: A) the Canadian prairies (zone 2 sub-zone 2), B) the central valley region in California, US (zone 5), C) the southern Great Plains in the US, (zone 7 sub-zone 4), D) the corn belt region of Heartland in the mid-western US (zone 8 sub-zone 1), E) southern Sonora in Mexico (zone 14), and F) the combined region of Chihuahua, Coahuila, and Nuevo Leon in Mexico (zone 15). The dashed vertical lines represent best merge thresholds. In panels A, B, C, D, and F, we selected minimum error rates with low fluctuations in error rate with these thresholds. In panel E, we selected the minimum error rate for all merge thresholds.

24

Table 5: List of merge thresholds for all RHSeg sub-zones.

Zone	Sub-zone	Merge threshold
1	1	20
2	1	18
2	2	17
3	1	18
3	2	14
3	3	12
3	4	20
4	1	15
5	1	11
6	1	9
6	2	10
6	3	13
6	4	12
6	5	10
7	1	19
7	2	17
7	3	27
7	4	16
7	5	21
8	1	21
8	2	29
8	3	18
9	1	12
9	2	7
9	3	15
10	1	18
10	2	17
11	1	9
12	1	15
13	1	16
13	2	8
14	1	14
15	1	17

The merge thresholds with the lowest error rate and lowest frequency of change in error rate across the merge thresholds was identified to be the best merge threshold for the sub-zone (Figure 12). The segmentation hierarchy corresponding to this best merge threshold had the most similarity with the sample field boundaries and the resulting object-based classification output was most likely the field boundaries on the ground. The RHSeg program was then used to derive the final segmentation output with this merge threshold for all tiles in the sub-zone (Table 5). This approach allowed automated derivation of the best segmentation hierarchy for the entire

sub-zone. The object-based classification was further refined by adding a minimum object size threshold of 9 pixels and a maximum of 2,000 pixels.
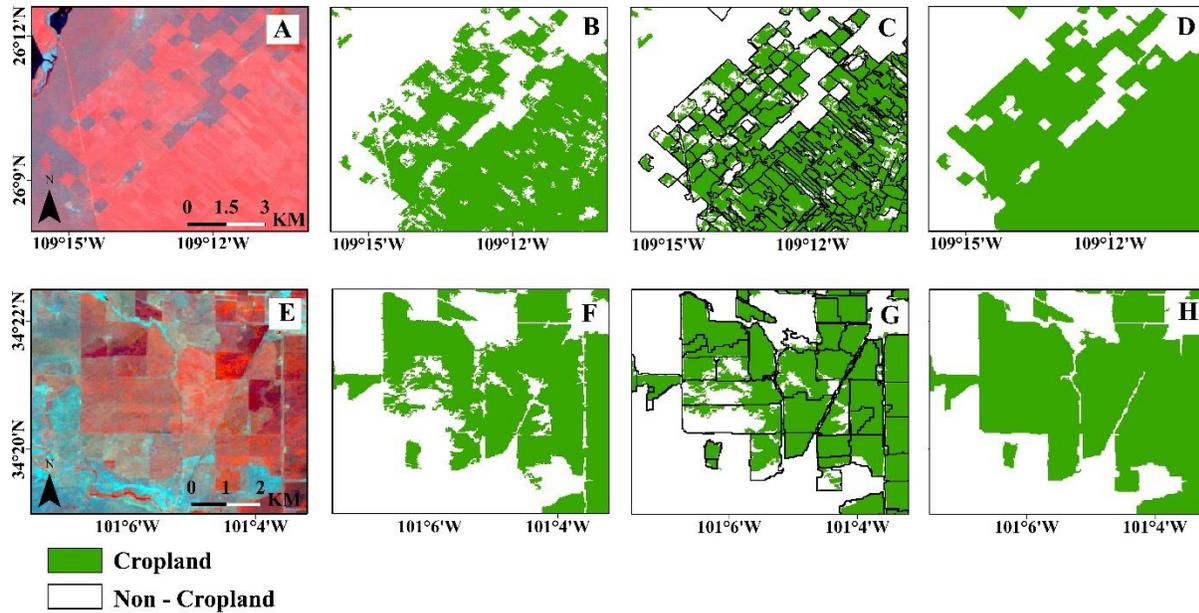
Figure 13: Landsat 5 TM data examples in west Sonora, Mexico (panel A) and northern Texas, US (panel E) were classified using the Random Forest classification (panel B & F), which were fused with RHSeg-derived maps of individual fields (panel C & G) to produce the final cropland extent maps (panels D and H). Each field was labeled as cropland or non-cropland (panel D & H) using a minimum cropland cover criterion of 25% of the pixels within each object.

The object-based classification output from RHSeg was intersected with the RF pixel-based classification output to fuse the map of objects with the pixel-based classification. Any object with less than 25% of its pixels as cropland in the fusion map was eliminated. The remaining objects within the cropland extent were labeled as cropland objects. All cropland objects were then merged to form a refined cropland extent for the tile which were further merged with other tiles to create the cropland extent for the zone. This approach greatly reduced the exclusions of individual crop fields and helped in removing small and scattered pixels labelled as cropland in non-cropland areas (Figure 13).

Object-based classification using RHSeg was performed only in zones that have SR data available. We did not perform object-based classification in zones 16 – 25. Instead, we performed a sieving and clumping operation using the *connectedPixelCount* function in GEE on the RF classification output. The *connectedPixelCount* function can be used to generate an image where each pixel contains the number of 4- or 8-connected neighbors. We used 2 as a threshold to sieve out smaller groups of pixels that have less than 2 4-connected neighboring pixels. The class labels for cropland and non-cropland pixels in the sieved output were 1 and 0, respectively.

A binary inversion was performed on this image where the cropland pixels were assigned label 0 and non-cropland label 1. The sieving operation was repeated to remove small groups of pixels resulting in smaller exclusions filled or clumped. Binary inversion was performed on the final output to assign the previous labels to the cropland and non-cropland classes as 1 and 0, respectively. The resulting output was the final cropland extent for the zone.

4. Programming and codes

The pixel-based RF classification was performed on Google Earth Engine (GEE) using Java Script application programming interface (API). The object-based classification using RHSeg was implemented using IDL on NAU Monsoon. The codes for the pixel-based classification, object-based classification, and their fusion are made available in a zip file and are available for download along with this ATBD.

5. Results

We produced a cropland extent map of North America for the nominal year 2010 at 30 m spatial resolution (Figure 14) by fusing the pixel-based classification and the object-based classification. Across the North American continent, a total cropland area 275.18 million hectares (Mha) was classified. The large regions of Canada, USA, and Mexico had cropland extent areas of 42.9 Mha, 169.2 Mha, and 35.9 Mha, respectively. The Central American countries had a total cropland area of 8.6 Mha, while the cropland extent for the Caribbean was 8.5 Mha.
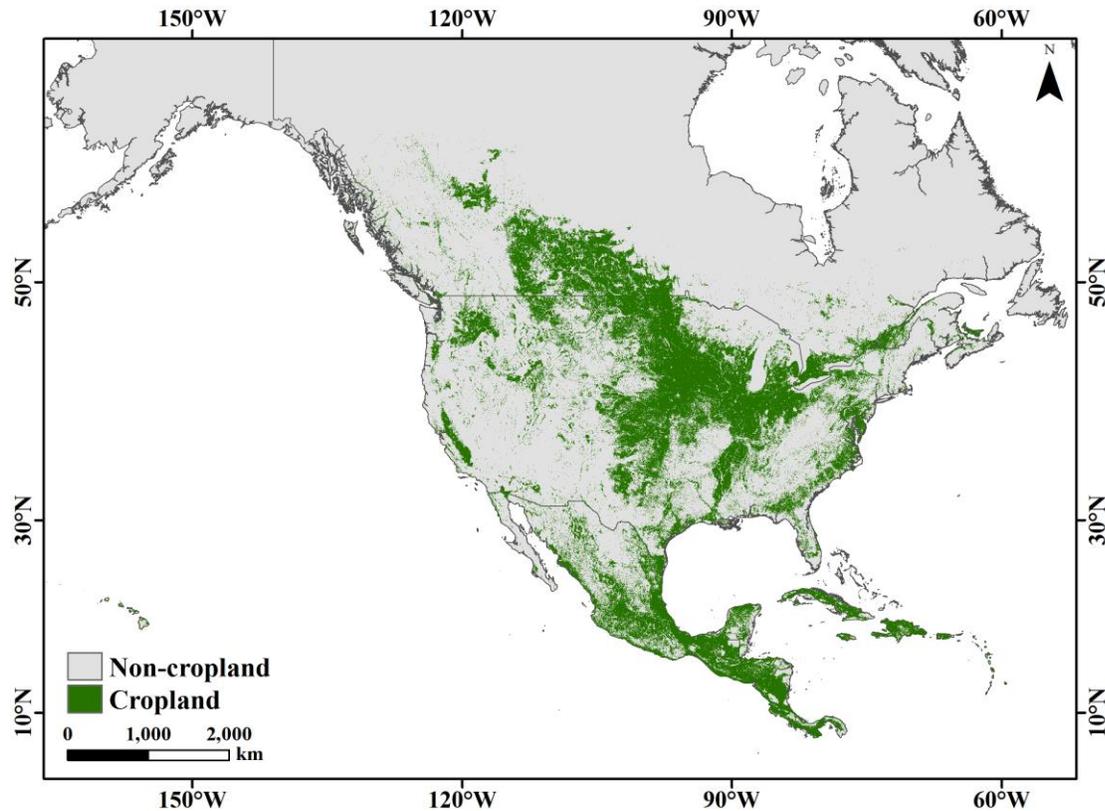
Figure 14: Cropland extent at 30m spatial resolution for the North American continent in the nominal year 2010.

27

6. North American cropland areas

We produced continental-scale cropland extent map of North America for the nominal year 2010 at 30 m spatial resolution by fusing the pixel-based classification using the RF classifier and the object-based classification from the RHSEG program. The country-level cropland areas in North America are listed in Table 6. Comparison of the cropland areas with the reported areas are shown in Figure 15. The total area for North American croplands was calculated to be 275.2 Mha.

Table 6: Country-wise net cropland area comparison between the nominal 2010 cropland extent and the reported cropland areas from multiple statistics agencies in North America.

Region	Cropland extent Area estimate 2010 (Ha)	Reported cropland Area 2010 (Ha)	Source agency
Anguilla	0	0	Food and Agriculture Organization Statistics
Antigua and Barbuda	0	0	Food and Agriculture Organization Statistics
Bahamas	11,150	11,998	Food and Agriculture Organization Statistics
Barbados	18,571	17,005	Food and Agriculture Organization Statistics
Belize*	385,466	157,000	Ministry of Agriculture, Belize
British Virgin Islands*	1,964	2,002	Food and Agriculture Organization Statistics
Canada	42,943,258	43,000,152	Agriculture and Agri-Food Canada
Costa Rica*	2,118,003	2,406,418	Ministry of Agriculture and Livestock, Costa Rica
Cuba*	4,218,071	4,221,000	National Office of Statistics and Information, Cuba
Dominica*	11,033	22,032	Food and Agriculture Organization Statistics
Dominican Republic*	2,033,977	2,065,281	Agricultural Sector Statistics, Dominican Republic
El Salvador*	1,030,295	883,666	Ministry of Agriculture, El Salvador
Grenada*	4,658	11,040	Food and Agriculture Organization Statistics
Guadeloupe*	65,863	24,080	Food and Agriculture Organization Statistics
Guatemala*	4,546,721	3,810,200	National Institute of Statistics, Guatemala
Haiti*	1,194,478	1,632,658	Food and Agriculture Organization Statistics
Hawaii	282,168	421,059	Department of Agriculture, State of Hawaii
Honduras*	3,354,977	3,201,512	National Institute of Statistics, Honduras
Jamaica*	290,929	270,368	Statistical Institute of Jamaica
Martinique	29,052	17,091	Food and Agriculture Organization Statistics
Mexico*	35,958,446	37,624,312	National Institute of Statistics and Geography, Mexico
Montserrat	330	2,001	Food and Agriculture Organization Statistics
Nicaragua*	4,922,544	3,732,512	Ministry of Agriculture and Forestry, Nicaragua
Panama*	2,133,809	2,051,119	National Institute of Statistics and Census, Panama
Puerto Rico*	355,040	170,407	United States Department of Agriculture
Saba	0	0	Food and Agriculture Organization Statistics
Saint Barthelemy	0	0	Food and Agriculture Organization Statistics
Saint Eustatius	0	0	Food and Agriculture Organization Statistics
Saint Kitts and Nevis	7,655	4,584	Food and Agriculture Organization Statistics
Saint Lucia	4,321	10,000	Food and Agriculture Organization Statistics
Saint Martin	0	0	Food and Agriculture Organization Statistics

Saint Vincent and the Grenadines	3,839	8,000	Food and Agriculture Organization Statistics
Sint Maarten	0	0	Food and Agriculture Organization Statistics
United States	169,233,515	166,244,617	United States Department of Agriculture
US Virgin Islands	5,768	2,000	United States Department of Agriculture

* Cropland area estimates and reported areas include cropland and managed pastures
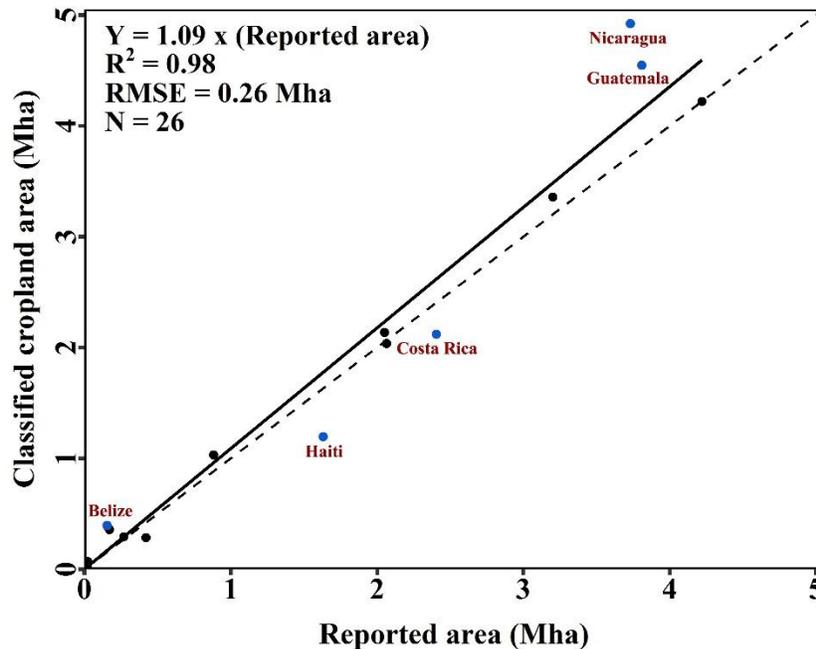
Figure 15: Cropland area estimates as compared with country level data from multiple agencies for the countries and regions in the North American continent in million hectares (Mha), with some of the outliers highlighted. This regression does not include Canada, the United States, and Mexico.

V.   Calibration Needs/Validation Activities

The cropland extent for the North American continent was assessed for its accuracy using three different methods: 1) Statistical accuracy assessment using randomly generated crop/non-crop samples, 2) Map-to-map comparison with available cropland maps, 3) Pixel-based and fusion-based classification output comparison, and 4) regression analysis with county and state cropland data, where available.

Statistical accuracy assessment of each AEZ for North America was performed by generating an error matrix for each zone (Congalton, 1991; Congalton and Green 2009). The error matrix was generated using a random sample of 250 sample units. Overall accuracy as well as producer's and user's accuracy (Story and Congalton, 1986) can be easily computed from the matrix.

Map-to-map comparison was performed for Canada, USA, and Mexico, wherein all pixels of the merged RF-RHSeg classification output were compared with the reference maps. We performed a map-to-map comparison for Canada using the nominal 2010 reference cropland extent derived from AAFC ACI for years 2009-2012. We similarly generated the nominal 2010 USDA CDL data by overlaying USDA CDL cropland extent from 2008-2012. In Mexico, the only spatial data available was the Agricultural Frontier series 2 boundaries of agricultural land use and irrigation digitized cropland boundaries for 2010-2011 (SIAP, 2017). We derived cropland extent from this map for the year 2010-2011. We then compared this reference cropland extent with our nominal year 2010 classification output for Mexico for map-to-map comparison.

In addition to the statistical accuracy assessment and map-to-map comparison, we compared the pixel-based output county areas with the fusion output county areas in Canada and the US. We also performed regression analysis for areas, where sub-country level cropland statistical reports were available and publicly accessible. The level of detail and availability of the statistical reports varied from country to country and we used the finest available level. For example, if both state and county level data were available, we used the county level statistical report. However, for many countries, only state level data were available. The regression analysis was performed using: 1) county cropland area statistics for the US, Canada, Mexico, Costa Rica, Hawaii, and Puerto Rico, and 2) state level cropland area statistics for Cuba, El Salvador, Dominican Republic, Jamaica, Nicaragua, and Panama. Table 1 lists all the data sources used for the county and state area statistics used in the regression analysis. FAO estimates of the cropland extent areas were compared with the cropland extent areas for the countries, where no spatial, county-wise, or state-wise cropland statistics could be obtained.

A. Accuracy assessment

Independent accuracy assessment was conducted for the 25 zones (Figure 2) in North America. For the entire North American continent, the overall accuracy was 93.4% with unweighted producer's accuracy at 85.4% (omission error of 14.6%) and unweighted user's accuracy at 74.5% (commission error of 25.5%) for the crop class (Table 7).

Across the 25 zones in the North American continent (Table 7):

1. Overall accuracies vary between 84.4-99.6%;
2. Producer's accuracies for the cropland class varies between 64.5-100% (errors of omissions: 0-35.5%); and
3. User's accuracies for the cropland class varies between 20-95% (errors of commissions (5-80%).

Across the 3 zones of Canada (Table 7):
1. Overall accuracies vary between 92.4-99.6%;
2. Producer's accuracies for the cropland class varies between 78.1-100% (errors of omissions: 0-21.9%); and
3. User's accuracies for the cropland class varies between 20-95% (errors of commissions (5-80%).

Across 9 zones of USA (Table 7):
1. Overall accuracies vary between 88.8-97.2%;
2. Producer's accuracies for the cropland class varies between 64.5-95.2% (errors of omissions: 4.8-35.5%); and
3. User's accuracies for the cropland class varies between 59.5-92.3% (errors of commissions (7.7-41.5%).

Across 6 zones of Mexico (Table 7):
1. Overall accuracies vary between 90.9-98.8%;
2. Producer's accuracies for the cropland class varies between 70-94% (errors of omissions: 6-30%); and
3. User's accuracies for the cropland class varies between 58.3-90.9% (errors of commissions (9.1-41.7%).

Across 5 zones of Central American and the Caribbean (Table 7):
1. Overall accuracies vary between 84.4-99.6%;
2. Producer's accuracies for the cropland class varies between 79.1-100% (errors of omissions: 0-20.9%); and
3. User's accuracies for the cropland class varies between 61.8-81.8% (errors of commissions (18.2-38.2%).

In this study our goal was to classify croplands with the lowest possible commission and omission errors across all the zones in North America. Using our classification approach, we were successful in achieving high producer's accuracies and consequently low omission errors for the cropland class for the majority of the 25 zones, that include zones in the US, Canada and Mexico. However, in some zones in Mexico, Central America and the Caribbean, the commission errors were high leading to low user's accuracy values.
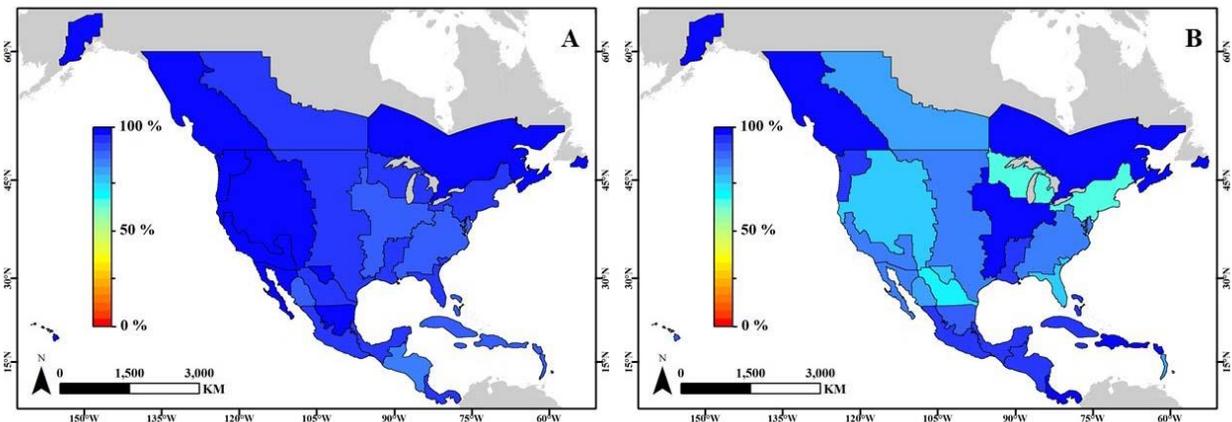
Figure 16: Spatial distribution of overall accuracy (panel A) and producer's accuracy for crop class (panel B) for all study area zones in North America.

31

Table 7: Statistical accuracy assessment of the fusion-based cropland extent map in all zones of the North American continent. The weighted accuracies are calculated using the percent zone areas of total land area in all zones as weights.

Zone	Region /country	Crop classified as crop	Crop classified as non-crop	Non-crop classified as crop	Non-crop classified as non-crop	Total crop in reference	Total non-crop in reference	Total crop in map	Total non-crop in map	Producer's accuracy for crop	Producer's accuracy for non-crop	User's accuracy for crop	User's accuracy for non-crop	Overall accuracy	% cropland area of North America total cropland area	% zone area of North America total area
Zone 1	Canada	1	0	4	245	1	249	5	245	100.0%	98.4%	20.0%	100.0%	98.4%	0.34%	6.28%
Zone 2	Canada	57	16	3	174	73	177	60	190	78.1%	98.3%	95.0%	91.6%	92.4%	12.27%	8.69%
Zone 3	Canada	10	0	1	239	10	240	11	239	100.0%	99.6%	90.9%	100.0%	99.6%	3.20%	12.22%
Zone 4	US	25	2	10	213	27	223	35	215	92.6%	95.5%	71.4%	99.1%	95.2%	1.19%	1.47%
Zone 5	US	24	5	2	219	29	221	26	224	82.8%	99.1%	92.3%	97.8%	97.2%	1.89%	3.09%
Zone 6	US	9	3	7	231	12	238	16	234	75.0%	97.1%	56.3%	98.7%	96.0%	4.23%	12.79%
Zone 7	US	63	14	10	163	77	173	73	177	81.8%	94.2%	86.3%	92.1%	90.4%	22.19%	12.73%
Zone 8	US	98	5	23	124	103	147	121	129	95.2%	84.4%	81.0%	96.1%	88.8%	17.70%	6.87%
Zone 9	US	20	11	9	210	31	219	29	221	64.5%	95.9%	69.0%	95.0%	92.0%	3.65%	5.83%
Zone 10	US	42	10	16	182	52	198	58	192	80.8%	91.9%	72.4%	94.8	89.6%	6.34%	5.78%
Zone 11	US	61	5	17	167	66	184	78	172	92.4%	90.8%	78.2%	97.1%	91.2%	3.04%	1.78%
Zone 12	US	22	9	15	204	31	219	37	204	71.0%	93.2%	59.5%	95.8%	90.4%	1.04%	1.33%
Zone 13	Mexico	10	2	1	236	12	237	11	238	83.3%	99.6%	90.9%	99.2%	98.8%	0.43%	1.86%
Zone 14	Mexico	53	14	8	173	67	181	61	187	79.1%	95.6%	86.9%	92.5%	91.1%	2.42%	2.08%
Zone 15	Mexico	14	6	10	217	20	227	24	223	70.0%	95.6%	58.3%	97.3%	93.5%	1.35%	2.51%
Zone 16	Mexico	52	6	16	167	58	183	68	173	89.7%	91.3%	76.5%	96.5%	90.9%	4.19%	2.89%
Zone 17	Mexico	22	3	4	220	25	224	26	223	88.0%	98.2%	84.6%	98.7%	97.2%	0.86%	1.20%
Zone 18	Mexico	47	3	13	166	50	179	60	169	94.0%	92.7%	78.3%	98.2%	93.0%	3.78%	2.17%
Zone 19	Central America	65	4	35	146	69	181	100	150	94.2%	80.7%	65.0%	97.3%	84.4%	5.57%	3.07%
Zone 20	Central America	52	0	23	171	52	194	75	171	100.0%	88.1%	69.3%	100.0%	90.7%	1.20%	1.05%
Zone 21	Caribbean	70	7	18	151	77	169	88	158	90.9%	89.4%	79.6%	95.6%	89.8%	1.64%	1.93%
Zone 22	Caribbean	71	3	32	142	74	174	103	145	96.0%	81.6%	68.9%	97.9%	85.9%	1.33%	1.20%
Zone 23	Caribbean	34	9	21	182	43	203	55	191	79.1%	89.7%	61.8%	95.3%	87.8%	0.05%	0.25%
Zone 24	Hawaii	36	4	8	202	40	210	44	206	90.0%	96.2%	81.8%	98.1%	95.2%	0.10%	0.11%
Zone 25	Alaska	4	0	1	245	4	246	5	245	100.0%	99.6%	80.0%	100.0%	99.6%	0.00%	0.82%
Weighted accuracies (by % zone areas)										85.4%	94.6%	74.5%		93.4%		

B. Map-to-map comparison

Our map-to-map comparison results indicate strong agreements for the regions with other existing cropland maps. Compared to the AAFC ACI map for Canada, the overall similarities of our map were 93.4% and 98.1% in zones 2 and 3, respectively (Table 8). In zone 2, which is the major agricultural zone across the Canadian states of Alberta, Saskatchewan, and Manitoba, the user's similarity of 82.7% and producer's similarity of 92.3% show low commission and omission errors. In zone 3, which includes agricultural area in the states of Ontario, Quebec, and New Brunswick, the omission error is moderately high with producer's similarity of crop at 62.4%.

The map-to-map comparison revealed an overall similarity >90% in all the US zones (Table 8). Zones 7, 8, and 10 consist of > 76% of the total cropland area in the US. Zone 8, the major corn and soybean growing region in the US, had the highest producer's similarity of 88% and the highest user's similarity of 94%. Zones 7 and 10 had producer's similarity of 86.2% and 78.8%, respectively, and user's similarities of 85.7% and 80.1%, respectively. While the overall similarity is mostly influenced by the large number of the non-cropland pixels, the high producer's and user's similarity for cropland show low commission and omission errors in these zones. In the US, zone 12 which includes croplands in Florida and Georgia had the lowest producer's similarity of 57% and user's similarity of 69.8%.

The map-to map comparison with the Mexico Agricultural frontiers 2010-2011 map revealed overall similarities > 86% (Table 8). The user's similarity ranges between 69% and 90%, and the producer's similarity ranges between 55% and 71%. The low user's similarity is associated with more croplands detected in our map compared to the Mexico Agricultural Frontiers map for 2010-2011. The spatial distribution of overall and producer's similarity for crop class is shown in Figure 17.
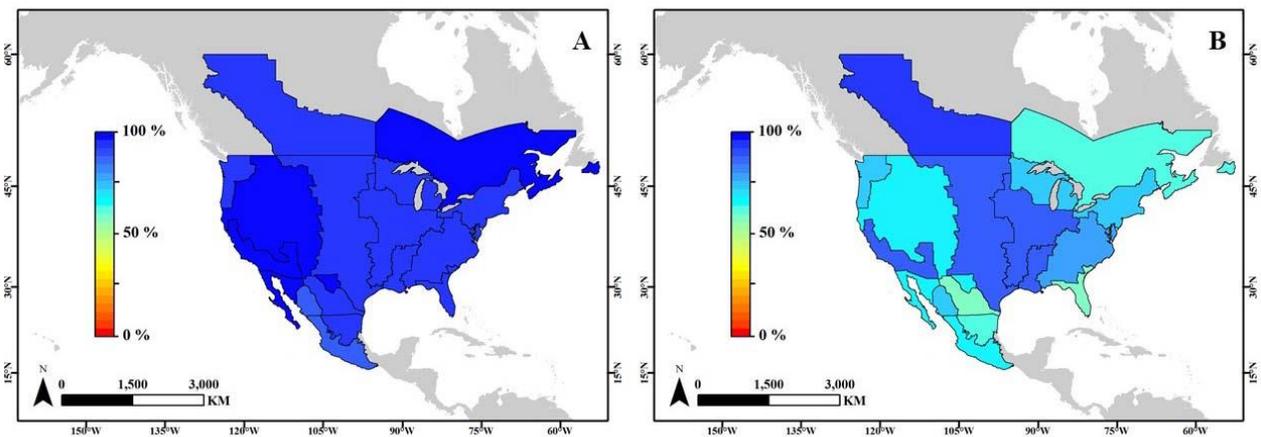
Figure 17: Spatial distribution of overall similarity (panel A) and producer's similarity for crop class (panel B) for study area zones with available reference data in North America.

33

Table 8: Similarities derived from map-to-map comparison of our nominal 2010 cropland extent with cropland extent derived from reference data in Canada, the US, and Mexico. The reference data sources are Agriculture and Agri-Food Canada (AAFC) annual crop inventory (ACI) for Canada, the United States Department of Agriculture (USDA) cropland data layer (CDL) for the US, and Servicio de Información Agroalimentaria y Pesquera (SIAP) agricultural frontier boundaries for Mexico.

Zone	Region/ country	Source	Producer's similarity for crop	Producer's similarity for non-crop	User's similarity for crop	User's similarity for non-crop	Overall similarity	% cropland area of North America total cropland area
Zone 2	Canada	AAFC ACI	92.3%	93.8%	82.7%	97.4%	93.4%	12.27%
Zone 3	Canada	AAFC ACI	62.4%	99.7%	91.5%	98.3%	98.1%	3.20%
Zone 4	US	USDA CDL	71.1%	99.0%	92.5%	95.2%	94.9%	1.19%
Zone 5	US	USDA CDL	88.1%	96.9%	78.5%	98.4%	95.9%	1.89%
Zone 6	US	USDA CDL	68.3%	98.7%	77.5%	97.9%	96.7%	4.23%
Zone 7	US	USDA CDL	86.2%	93.2%	85.7%	93.5%	90.9%	22.19%
Zone 8	US	USDA CDL	88.7%	95.0%	94.1%	90.3%	92.0%	17.70%
Zone 9	US	USDA CDL	74.8%	94.8%	64.4%	96.7%	92.5%	3.65%
Zone 10	US	USDA CDL	78.8%	95.2%	80.1%	94.9%	92.0%	6.34%
Zone 11	US	USDA CDL	85.7%	93.4%	85.3%	93.5%	91.0%	3.04%
Zone 12	US	USDA CDL	57.0%	96.0%	69.8%	93.2%	90.5%	1.04%
Zone 13	Mexico	SIAP	65.6%	99.3%	79.8%	98.5%	97.9%	0.43%
Zone 14	Mexico	SIAP	71.2%	92.1%	69.9%	92.5%	87.8%	2.42%
Zone 15	Mexico	SIAP	55.6%	98.6%	80.9%	95.4%	94.4%	1.35%
Zone 16	Mexico	SIAP	66.2%	93.8%	78.5%	89.1%	86.8%	4.19%
Zone 17	Mexico	SIAP	64.1%	99.0%	90.9%	94.9%	94.5%	0.86%

C. Pixel-based and fusion-based classification output comparison

Our comparison of the pixel-based cropland classification from the random forest classifier with the county area estimates reported in the USDA CDL for the US indicate consistent underestimation of cropland areas (Figure 18). The R^2 value of 0.96 indicates high correlation, but the slope of 0.83 indicates a general trend of underestimation in comparison with USDA CDL. This trend is also seen across the areas derived using the pixel-based random forest classification in the agricultural census sub-divisions in Canada, when we compare our results to the AAFC ACI estimates. While the R^2 value of 0.96 indicates good correlation, the slope of 0.71 shows that the pixel-based classification underestimates the cropland areas.

In contrast, the classification output from the fusion of pixel-based and object-based classifications showed results more consistent with the reference maps. The R^2 value of 0.97 and a slope of 1.01 between our cropland area estimates and the USDA CDL indicates that the fusion-based cropland extent map accurately estimates the county areas. In Canada, the R^2 value of 0.98 and slope of 0.86 indicate comparatively accurate agricultural census sub-division area estimates in contrast with the pixel-based output. We, therefore, selected the fusion-based output for North America as our final output for the nominal year 2010.
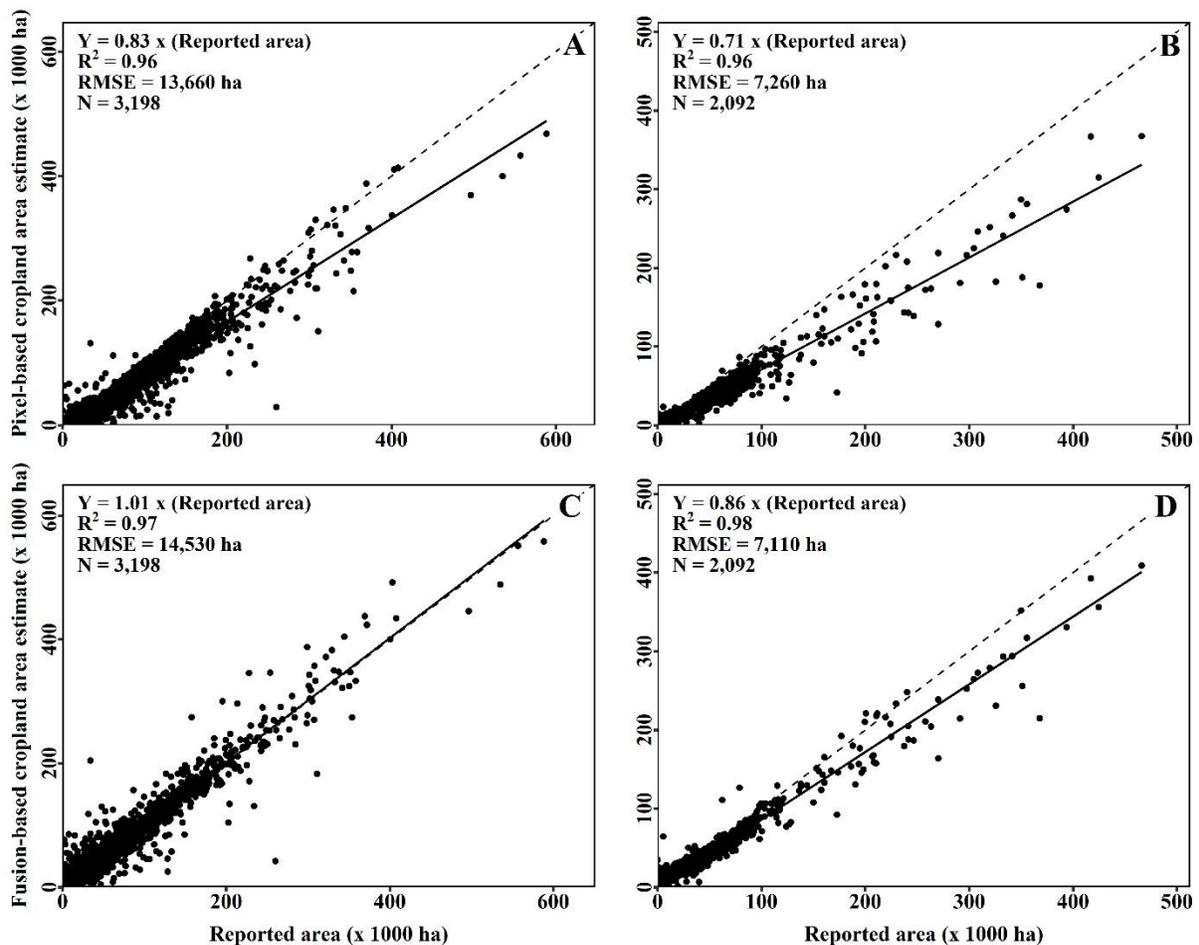
Figure 18: Cropland area (ha) estimates using pixel-based random forest classification compared with the USDA CDL cropland extent at county level in the US (panel A) and the Canadian

AAFC ACI (panel B), and the fusion-based classification map compared with the USDA CDL cropland extent at county level in the US (panel C) and the Canadian AAFC ACI cropland extent at agricultural census sub-division level (panel D).

D. Regression analysis

Our regression analysis with county level data included four regions: 1) US, 2) Canada, 3) Mexico, and 4) Costa Rica (Figure 18 panels C & D; Figure 19). The regression analysis indicated that the classified cropland extent is consistent with the available reference maps in terms of areas of smaller sub-divisions for the North American continent. In the US, the coefficient of determination (R^2) between the classified cropland areas and the USDA CDL cropland extent was 0.97. In addition to a strong correlation, the slope of 1.01 from this regression model suggests a strong overall agreement in the cropland area estimates. The root mean squared error (RMSE) value for this regression was 14,530 hectares (ha) (Figure 18 panel C). In Canada, the R^2 value between the classified cropland area and the AAFC ACI cropland extent was 0.97 at agricultural census sub-division level. The slope of 0.86 indicated a slight under-classification of croplands in comparison to AAFC ACI estimates. The distribution of points around the regression line is compact with a low RMSE of 7,110 Ha (Figure 15 panel D).

Mexico classification output had a R^2 value of 0.84 when compared with SIAP Agricultural Frontier boundaries cropland extent at the district level (Figure 19). The relatively large RMSE of 15,023 ha suggests that the regression relationship is weaker for Mexico, while the slope of 1.09 suggests a slight over-classification. In Costa Rica, we compared district level reported areas from 2014 with our cropland extent map. This regression was strong with an R^2 value of 0.95 and a low RMSE of 8,974 ha. The slope of 0.87 indicated under-classification, which might be partially due to the temporal difference in the classification output and the reference.

Additionally, we compared state-level cropland areas with the classified cropland extent for 1) Nicaragua, 2) Dominican Republic, 3) Panama, 4) Jamaica, 5) El Salvador, and 6) Cuba (Figure 19). The state level comparisons generally showed moderately strong relationships with the classified cropland extents. In Nicaragua, a R^2 of 0.91 with a slope of 1.19 shows a strong correlation with slight over-classification. While the Dominican Republic had a wider spread around the regression line with R^2 of 0.73, a slope value of 0.98 represents mapped areas very close to the actual cropland areas. Panama had a high R^2 value of 0.94 and slope of 1.04 which shows strong correlation with reported cropland areas. The R^2 value for Jamaica is 0.89 and its slope of 0.92 show under-classification. Cuba with its R^2 value of 0.96 and regression slope of 1.12 has small amount of over-classification. El Salvador had the lowest correlation of 0.8 and slope of 1.68, which indicates large over-classification. In addition, the large RMSE for El Salvador indicates a comparatively weaker performance of our classification model for this country (Figure 19).
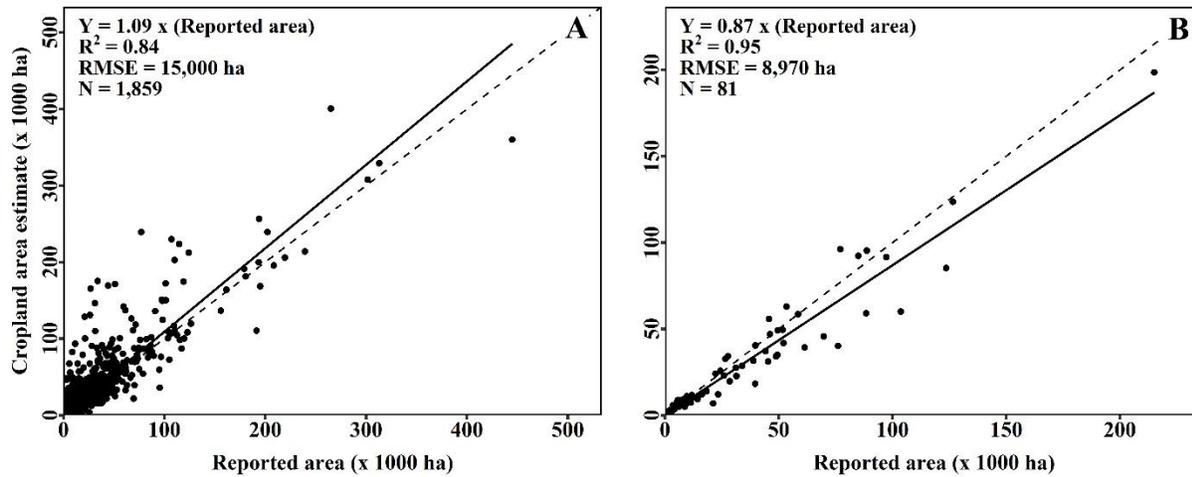
Figure 19: Cropland area (ha) estimates compared with: A) Servicio de Información Agroalimentaria y Pesquera (SIAP) Agricultural Frontiers cropland extent for 2010-2011 at district level, and B) Costa Rican Ministry of Agriculture census 2014 total farm area at district level.
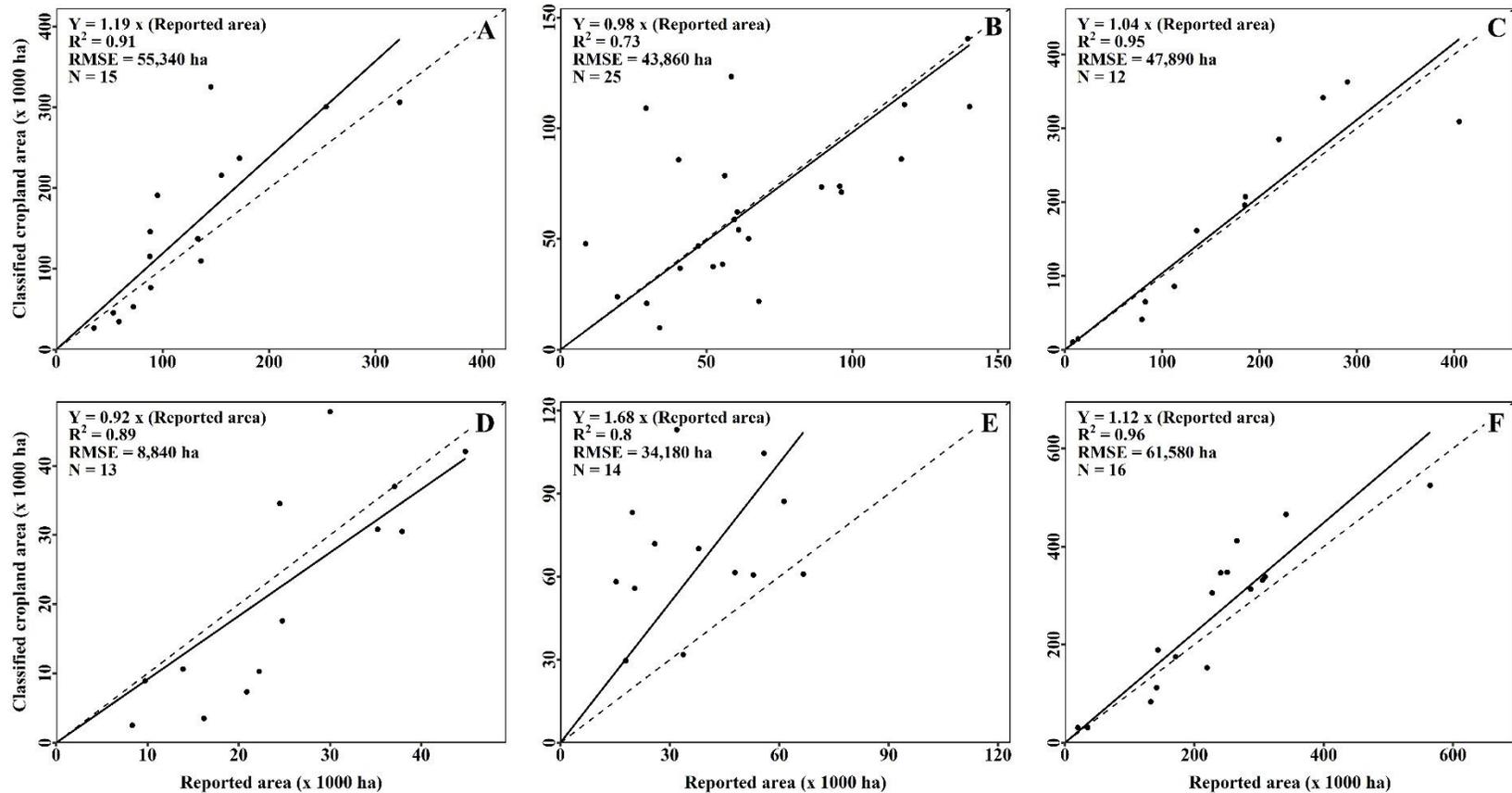
Figure 20: Cropland areas (ha) estimates compared with state level data on: A) total cropland area in Nicaragua's agricultural census 2012, National Institute for Development Information, B) total cropland area from 2015 agricultural census by Agricultural Sector Statistics, Dominican Republic, C) cropland area derived from 2010 land use statistics by National Institute of Statistics, Panama, D) farm area in 2007 from Statistical Institute of Jamaica, E) agricultural area for 2008 from the Ministry of Agriculture and Livestock of El Salvador, and F) cropland area for 2015 obtained from National Bureau of Statistics, Cuba. The total cropland area includes croplands and managed pastures.

VI.   Constraints and limitations

The overall accuracies of the cropland extent map across all zones in the US is > 90%. The regression analysis of county statistics for our cropland extent with those derived from USDA CDL show strong agreement with slope of 1.01, R^2 of 0.97, and RMSE of 14,674 ha. Some of the overclassified counties are found in zones 4, 8, 9, and 12, where some of the wetlands, pastures, and roads were classified as croplands. The counties with large omission errors occur in zones 6 and 12, and southern parts of zone 5 due to orchards being classified as non-croplands and exclusion of fallow cropland. In the US, the validation results may be biased towards USDA CDL since it was our only independent data source. The accuracies of our cropland extent may also reflect commission and omission errors in USDA CDL. However, the high overall accuracies (> 90%) of the USDA CDL suggest that this bias effect may be low (Howard and Wylie, 2014).

In Canada, the map-to-map comparison of our classification output with the reference cropland extent derived from AAFC ACI shows high overall accuracies. The regression analysis yields a slope of 0.86, $R^2 = 0.97$, and a root mean squared error of 7,469 ha, indicating omission errors in comparison to the AAFC ACI. The zones that have the most commission are 1 and 3. Zone 1 is not classified in AAFC ACI 2009-2012. Zone 2 shows omission errors in comparison to AAFC ACI, which occur in croplands near Calgary, and along the eastern coast of Lake Manitoba, for example. The commissions in our map in zone 3 occur in the districts of southern Ontario and in the croplands along the St. Lawrence River. Upon visual inspection of AAFC ACI-derived cropland extent with high-resolution background imagery in GEE, we note that AAFC ACI appears to have commission errors in some zones, while the reported overall accuracies are > 85%. This study did not quantitatively assess the commission errors of AAFC ACI.

The Mexico agricultural frontiers 2 map (SIAP, 2017) is hand digitized and has inclusions as well as exclusions in many parts of southern Mexico. The agricultural frontiers 2 map does not cover the Yucatan peninsula in southern Mexico. Our map-to-map comparison and district-wise regression analysis were, therefore, restricted only in the zones that include the agricultural frontiers 2 map, although our cropland extent map included all croplands in Mexico. The R^2 value of 0.84 suggests good agreement, but also indicates some disagreements. In particular, these disagreements occur as omission errors in zones 15 and 16 or as commission errors in the Yucatan peninsula in zone 15. These differences are due to small field sizes, highly wet climate, and low temporal density of Landsat data in southern Mexico.

In our classification of Central America, Mexico, the Caribbean Islands, and Hawaii, we included cropland pasture in the cropland extent. This is because a majority of the farms in Central America, Mexico, and the Caribbean Islands are small and could not be detected due to: a) Landsat surface reflectance data being limited and only TOA data were available, and b) the small farms often rotate between croplands and pasture, and c) slash and burn agriculture was common further challenging the cropland distinction.

Spatial data or land cover maps are rare for Central America and the Caribbean Islands. The agricultural statistics available in some of the countries were, therefore, essential in validating our cropland extent classification in those regions in addition to statistical accuracy assessment. Low accuracies in the statistical assessment and low R^2 values in the regression analysis with agricultural statistics from some countries, such as Dominican Republic and El

Salvador, are related to the: 1) limited data quality due to common presence of clouds, 2) high rainfall, and 3) low density of images in the time-series data.

In the object-based classification, the identification and selection of appropriate field size was essential. Field sizes varied across the North American continent and among the different zones. The same parameters, therefore, could not be used in the object-based classification of the entire continent. The optimal merge threshold was similarly different in each sub-zone for the RHSEG object-based classification. We implemented the object-based RHSeg classification only in the regions and zones with sufficient Landsat SR data availability: the US, Canada, and the northern parts of Mexico. RHSeg was not implemented in southern Mexico, Central American region, and the Caribbean Islands due to: 1) low data quality of Landsat 5 TM TOA with clouds, 2) small holder farms with only small differences from surrounding vegetation, and 3) inability to form a seamless mosaic using data from two periods in a year. Instead, we used sieving and clumping to reduce the noise in the cropland extent classification in southern Mexico, Central American region, and the Caribbean Islands.

The fusion reduced within-field variability and removed incorrectly classified small clusters of cropland pixels. The use of a threshold of 25% cropland pixels was critical in the fusion with the pixel-based classification. The fusion also resulted in some omission and commission errors, when multiple crop fields were merged together during the object-based classification to form one object since crop fields commonly occurred adjacent to each other. Our regression analysis (Figure 15) showed that the pixel-based classification for North America consistently underestimated cropland areas in the US counties and Canadian agricultural census sub-divisions. In contrast, the fusion-based classification significantly improved the output as indicated by comparing the county areas in the US and agricultural census sub-divisions in Canada with those derived from USDA CDL and AAFC ACI, respectively.

Other limitations in the North American cropland extent for nominal year 2010 are largely due to the data availability across the North American continent (Figure 3). Regions of significant cloud cover, such as the Central American countries, can have large variability in the Landsat 5 TM and Landsat 7 ETM+ image collection, which cause errors in the RF classification. Lastly, regions with relatively small field sizes, large amount of rainfall, irregular patterns of irrigation, and longer growing periods may have lower accuracies due to highly similar time-series curves between crop and non-crop samples.

VII. Acknowledgements

VIII. Publications

The following publications are related to the development of the above croplands products:

1. Peer-reviewed publications specific to this study

Massey, R., Sankey, T.T., Yadav, K., Congalton, R.G., Tilton, J.C., Thenkabail, P.S., (2017). Landsat-derived North American continental-scale cropland extent classification using high performance computing platforms. In preparation.

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.

2. Peer-reviewed publications within GFSAD project

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).

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.

3. 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)

4. Other relevant past publications prior to GFSAD project

Biggs, T., Thenkabail, P.S., Krishna, M., GangadharaRao Rao, P., and Turral, H., 2006. Vegetation phenology and irrigated area mapping using combined MODIS time-series, ground surveys, and agricultural census data in Krishna River Basin, India. International Journal of Remote Sensing. 27(19):4245-4266.

Biradar, C.M., Thenkabail, P.S., Noojipady, P., Yuanjie, L., Dheeravath, V., Velpuri, M., Turral, H., Gumma, M.K., Reddy, O.G.P., Xueliang, L. C., Schull, M.A., Alankara, R.D., Gunasinghe, S., Mohideen, S., Xiao, X. 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(2). 114-129. doi:10.1016/j.jag.2008.11.002. January, 2009.

Dheeravath, V., Thenkabail, P.S., Chandrakantha, G, Noojipady, P., Biradar, C.B., Turral. H., Gumma, M.1, Reddy, G.P.O., Velpuri, M. 2010. Irrigated areas of India derived using MODIS 500m data for years 2001-2003. ISPRS Journal of Photogrammetry and Remote Sensing. http://dx.doi.org/10.1016/j.isprsjprs.2009.08.004. 65(1): 42-59.

Thenkabail, P.S. 2012. Special Issue Foreword. Global Croplands special issue for the August 2012 special issue for Photogrammetric Engineering and Remote Sensing. PE&RS. 78(8): 787- 788.Thenkabail, P.S. 2012. Guest Editor for Global Croplands Special Issue. Photogrammetric Engineering and Remote Sensing. PE&RS. 78(8).

Thenkabail, P.S., Biradar C.M., Noojipady, P., Cai, X.L., Dheeravath, V., Li, Y.J., Velpuri, M., Gumma, M., Pandey, S. 2007a. Sub-pixel irrigated area calculation methods. Sensors Journal (special issue: Remote Sensing of Natural Resources and the Environment (Remote Sensing SensorsEdited by Assefa M. Melesse). 7:2519-2538. http://www.mdpi.org/sensors/papers/s7112519.pdf.

Thenkabail, P.S., Biradar C.M., Noojipady, P., Dheeravath, V., Li, Y.J., Velpuri, M., Gumma, M., Reddy, G.P.O., Turral, H., Cai, X. L., Vithanage, J., Schull, M., and Dutta, R. 2009a. Global irrigated area map (GIAM), derived from remote sensing, for the end of the last millennium. International Journal of Remote Sensing. 30(14): 3679-3733. July, 20, 2009.

Thenkabail, P.S., Biradar, C.M., Turral, H., Noojipady, P., Li, Y.J., Vithanage, J., Dheeravath, V., Velpuri, M., Schull M., Cai, X. L., Dutta, R. 2006. An Irrigated Area Map of the World (1999) derived from Remote Sensing. Research Report # 105. International Water Management Institute. Pp. 74. Also, see under documents in: http://www.iwmigiam.org.

Thenkabail, P. S.; Dheeravath, V.; Biradar, C. M.; Gangalakunta, O. P.; Noojipady, P.; Gurappa, C.; Velpuri, M.; Gumma, M.; Li, Y. 2009b. Irrigated Area Maps and Statistics of India Using Remote Sensing and National Statistics. Journal Remote Sensing. 1:50-67. http://www.mdpi.com/2072-4292/1/2/50.

Thenkabail, P.S., GangadharaRao, P., Biggs, T., Krishna, M., and Turral, H., 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(9): 1029-1040. (Second Place Recipients of the 2008 John I. Davidson ASPRS President's Award for Practical papers).

Thenkabail, P.S., Hanjra, M.A., Dheeravath, V., Gumma, M.K. 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 open access journal. 2(1):211-261. doi:10.3390/rs2010211. http://www.mdpi.com/2072-4292/2/1/211

Thenkabail P.S., Knox J.W., Ozdogan, M., Gumma, M.K., Congalton, R.G., Wu, Z., Milesi, C., Finkral, A., Marshall, M., Mariotto, I., You, S. Giri, C. and Nagler, P. 2012. Assessing future risks to agricultural productivity, water resources and food security: how can remote sensing help? Photogrammetric

Engineering and Remote Sensing, August 2012 Special Issue on Global Croplands: Highlight Article. 78(8): 773-782.

Thenkabail, P.S., Schull, M., Turral, H. 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. Remote Sensing of Environment, 95(3): 317-341.

Velpuri, M., Thenkabail, P.S., Gumma, M.K., Biradar, C.B., Dheeravath, V., Noojipady, P., Yuanjie, L.,2009. Influence of Resolution or Scale in Irrigated Area Mapping and Area Estimations. Photogrammetric Engineering and Remote Sensing (PE&RS). 75(12): December 2009 issue.

5. Books and Book Chapters

Teluguntla, P., Thenkabail, P.S., Xiong, J., Gumma, M.K., Giri, C., Milesi, C., Ozdogan, M., Congalton, R.,Tilton, J.,Sankey,T.R., Massey, R., Phalke, A., and Yadav, K. 2015. Global Food Security Support Analysis Data at Nominal 1 km (GFSAD1 km) Derived from Remote Sensing in Support of Food Security in the Twenty-First Century: Current Achievements and Future Possibilities, Chapter 6. In Thenkabail, P.S., (Editor-in-Chief), 2015. "Remote Sensing Handbook" (Volume II): Land Resources Monitoring, Modeling, and Mapping with Remote Sensing. Taylor and Francis Inc. Press, Boca Raton, London, New York. ISBN 9781482217957 - CAT# K22130. Pp. 131-160

Biradar, C.M., Thenkabail. P.S., Noojipady, P., Li, Y.J., Dheeravath, V., Velpuri, M., Turral, H., Cai, X.L., Gumma, M., Gangalakunta, O.R.P., Schull, M., Alankara, R.D., Gunasinghe, S., and Xiao, X. 2009. Book Chapter 15: Global map of rainfed cropland areas (GMRCA) and stastistics using remote sensing. Pp. 357-392. In the book entitled: "Remote Sensing of Global Croplands for Food Security" (CRC Press- Taylor and Francis group, Boca Raton, London, New York. Pp. 475. Published in June, 2009. (Editors: Thenkabail. P., Lyon, G.J., Biradar, C.M., and Turral, H.).

Gangalakunta, O.R.P., Dheeravath, V., Thenkabail, P.S., Chandrakantha, G., Biradar, C.M., Noojipady, P., Velpuri, M., and Kumar, M.A. 2009. Book Chapter 5: Irrigated areas of India derived from satellite sensors and national statistics: A way forward from GIAM experience. Pp. 139-176. In the book entitled: "Remote Sensing of Global Croplands for Food Security" (CRC Press- Taylor and Francis group, Boca Raton, London, New York. Pp. 475. Published in June, 2009. (Editors: Thenkabail. P., Lyon, G.J., Biradar, C.M., and Turral, H.).

Li, Y.J., Thenkabail, P.S., Biradar, C.M., Noojipady, P., Dheeravath, V., Velpuri, M., Gangalakunta, O.R., Cai, X.L. 2009. Book Chapter 2: A history of irrigated areas of the world. Pp. 13-40. In the book entitled: "Remote Sensing of Global Croplands for Food Security" (CRC Press- Taylor and Francis group, Boca Raton, London, New York. Pp. 475. Published in June, 2009. (Editors: Thenkabail. P.,Lyon, G.J., Biradar, C.M., and Turral, H.).

Thenkabail, P.S., Lyon, G.J., and Huete, A. 2011. Book Chapter # 1: Advances in Hyperspectral Remote Sensing of Vegetation. In Book entitled: "Remote Sensing of Global Croplands for Food Security" (CRC Press- Taylor and Francis group, Boca Raton, London, New York. Edited by Thenkabail, P.S., Lyon, G.J., and Huete, A. Pp. 3-38.

Thenkabail. P.S., Biradar, C.M., Noojipady, P., Dheeravath, V., Gumma, M., Li, Y.J., Velpuri, M., Gangalakunta, O.R.P. 2009c. Book Chapter 3: Global irrigated area maps (GIAM) and statistics using remote sensing. Pp. 41-120. In the book entitled: "Remote Sensing of Global Croplands for Food Security" (CRC Press- Taylor and Francis group, Boca Raton, London, New York. Pp. 475. Published in June, 2009. (Editors: Thenkabail. P., Lyon, G.J., Biradar, C.M., and Turral, H.).

Thenkabail. P., Lyon, G.J., Turral, H., and Biradar, C.M. (Editors) 2009d. Book entitled: "Remote Sensing of Global Croplands for Food Security" (CRC Press- Taylor and Francis group, Boca Raton, London, New York. Pp. 556 (48 pages in color). Published in June, 2009. Reviews of this book: http://www.crcpress.com/product/isbn/9781420090093 http://gfmt.blogspot.com/2011/05/review-remote-sensing-of-global.html

Thenkabail, P.S. and Lyon, J.G. 2009. Book Chapter 20: Remote sensing of global croplands for food security: way forward. Pp. 461-466. In the book entitled: "Remote Sensing of Global Croplands for Food Security" (CRC Press- Taylor and Francis group, Boca Raton, London, New York. Pp. 475. Published in June, 2009. (Editors: Thenkabail. P., Lyon, G.J., Biradar, C.M., and Turral, H.).

Turral, H., Thenkabail, P.S., Lyon, J.G., and Biradar, C.M. 2009. Book Chapter 1: Context, need: The need and scope for mapping global irrigated and rain-fed areas. Pp. 3-12. In the book entitled: "Remote Sensing of Global Croplands for Food Security" (CRC Press- Taylor and Francis group, Boca Raton, London, New York. Pp. 475. Published in June, 2009. (Editors: Thenkabail. P., Lyon, G.J., Biradar, C.M., and Turral, H.).

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

For North America cropland extent, feel free to write to:
Richard Massey rm885@nau.edu
Temuulen Sankey Temuulen.Sankey@nau.edu

More details about the GFSAD project and products can be found at: globalcroplands.org

X.   Citations

Massey, R., Sankey, T.T., Yadav, K., Congalton, R.G., Tilton, J.C., Thenkabail, P.S. (2017). *NASA Making Earth System Data Records for Use in Research Environments (MEaSUREs) Global Food Security-support Analysis Data (GFSAD) Cropland Extent 2010 North America 30*

m V001 [Data set]. NASA EOSDIS Land Processes DAAC. doi:
10.5067/MEaSUREs/GFSAD/GFSAD30NACE.001

XI.   References

Boryan, C., Yang, Z., Mueller, R., Craig, M., 2011. Monitoring US agriculture: the US Department of Agriculture, National Agricultural Statistics Service, Cropland Data Layer Program. Geocarto Int. 26, 341–358.

Boryan, C.G., Zhengwei Yang, 2014. Implementation of a new automatic stratification method using geospatial cropland data layers in NASS area frame construction. 2014 IEEE Geosci. Remote Sens. Symp. 2110–2113.

Breiman, L., 2001. Random forests. Mach. Learning 45, 5–32
Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A., 1984. Classification and Regression Trees, first ed. Chapman and Hall/CRC, Belmont, CA

Brown, J.F., & Pervez, M.S., 2014. Merging remote sensing data and national agricul- tural statistics tomodel change in irrigated agriculture. Agricultural Systems, 127, 28–40.

Brown, J.F., Maxwell, S., Pervez, M.S., 2009. Mapping irrigated lands across the United States using MODIS satellite imagery. In: Thenkabail, P.S., Lyon, J.G., Turral, H., Biradar, C.M. (Eds.), Remote Sensing of Global Croplands for Food Security. Taylor & Francis, Boca Raton, Florida, USA, pp. 177–198.

Cerquiglini, C., Claro, J., Giusti, A.M., Karumathy, G., Mancini, D., Marocco, E., Mascianá, P., Michetti, M., Milo, M., 2016. Food Outlook June 2016. Food Agric. Organ. United Nations 14.

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

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

Dong, J., Xiao, X., Menarguez, M.A., Zhang, G., Qin, Y., Thau, D., Biradar, C., Moore, B., 2015. Mapping paddy rice planting area in northeastern Asia with Landsat 8 images, phenology-based algorithm and Google Earth Engine. Remote Sens. Environ. 185, 142-154.

Economic Research Service. 2000. USDA Ag. Info. Bulletin No. 760.

Erickson, T., 2014. Multi-source geospatial data analysis with Google Earth Engine. American Geophysical Union, Fall Meeting 2014, abstract #IN53E-05.

Farr, T.G., Rosen, P.A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., et al. (2007). The shuttle radar topography mission. Reviews of Geophysics, 45.

Fischer, G., van Velthuizen, H., Nachtergaele, F.O., 2000. Global Agroecological Zones Assessment: Methodology and Results (IIASA Interim Report IR-00-064).

Fisette, T., A. Davidson, B. Daneshfar, P. Rollin, and L. Campbell, 2014. Annual Space-Based Crop Inventory for Canada: 2009-2014. Geoscience and Remote Sensing Symposium (IGARSS), 2014 IEEE International, Quebec City, QC, pp. 5095–5098.

Fisette, T., P. Rollin, Z. Aly, L. Campbell, B. Daneshfar, P. Filyer, A. Smith, A. Davidson, J. Shang, and I. Jarvis, 2013. AAFC Annual Crop Inventory: Status and Challenges. The 2nd International Conference on Agro-Geoinformatics, Fairfax, VA, pp. 270–274.

Foley, J.A., Ramankutty, N., Brauman, K.A., Cassidy, E.S., Gerber, J.S., Johnston, M., Mueller, N.D., O'Connell, C., Ray, D.K., West, P.C., Balzer, C., Bennett, E.M., Carpenter, S.R., Hill, J., Monfreda, C., Polasky, S., Rockström, J., Sheehan, J., Siebert, S., Tilman, D., Zaks, D.P.M., O'Connell, C., 2011. Solutions for a cultivated planet. Nature 478, 337–42.

Friedl, M., & Brodley, C., 1997. Decision tree classification of land cover from remotely sensed data. Remote Sensing of Environment, 61(3), 399−409.

Gao, B., 1996. NDWI — A normalized difference water index for remote sensing of vegetation liquid water from space. Remote Sensing of Environment, 58, 257–266.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sens. Environ.

Han, W., Yang, Z., Di, L., Mueller, R., 2012. CropScape: a Web service based application for exploring and disseminating US conterminous geospatial cropland data products for decision support. Comput. Electron. Agric. 84, 111–123.

Hancher, M., 2013. Planetary-scale geospatial data Analysis techniques in Google's Earth Engine platform. American Geophysical Union, Fall Meeting, abstract #IN52A-07

Hansen, M., Dubayah, R., Defries, R., 1996. Classification trees: an alternative to traditional land cover classifiers. International Journal of Remote Sensing 17 (5), 1075–1081

Homer, C., Dewitz, J., Fry, J., Coan, M., Hossain, N., Larson, C., Herold, N., McKerrow, A., VanDriel, N., Wickham, J., 2007. Completion of the 2001 National Land Cover Database for the counterminous United States. Photogramm. Eng. Remote. Sens. 73 (4), 337.

Homer, C., Huang, C., Yang, L., Wylie, B., Michael, C., 2004. Development of a 2001 National Land Cover Database for the United States. Photogramm. Eng. Remote. Sens. 70 (7), 829–840.

Homer, C.H., Fry, J.A., Barnes, C.A., 2012. The National Land Cover Database. U. S. Geol. Surv. Fact Sheet 3020 (4), 1–4.

Housman, I.; Tanpipat, V.; Biswas, T.; Clark, A.; Stephen, P.; Maus, P.; Megown, K., 2015. Monitoring forest change in Southeast Asia: case studies for USAID Lowering Emissions in Asia's Forests. RSAC-10108-RPT1. Salt Lake City, UT: U.S. Department of Agriculture, Forest Service, Remote Sensing Applications Center, 16.

Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., & Ferreira, L. G., 2002. Overview of the Radiometric and Biophysical Performance of the MODIS Vegetation Indices. Remote Sensing of Environment, 83, 195–213

Inglada, J., Arias, M., Tardy, B., Hagolle, O., Valero, S., Morin, D., Dedieu, G., Sepulcre, G., Bontemps, S., Defourny, P., et al., 2015. Assessment of an operational system for crop type map production using high temporal and spatial resolution satellite optical imagery. Remote Sensing 7, 12356–12379.
Instituto Nacional de Estadística y Geografía (INEGI), 2007. Digital map of Mexico. http://gaia.inegi.org.mx/mdm5/viewer.html. Agricultural Atlas of Mexico: Agricultural Census 2007. Last accessed 1 Jul 2017.

Johnson, D.M., Mueller, R., 2010. The 2009 cropland data layer. Photogramm. Eng. Remote Sensing 1202–1205.

Key, C. H., & Benson, N. (1999). The Normalized Burn Ratio (NBR): A Landsat TM radiometric measure of burn severity. www.nrmsc.usgs. gov/research/ndbr.htm.

Lawrence, R. L., Wood, S. D., & Sheley, R. L., 2006. Mapping invasive plants using hyperspectral imagery and Breiman Cutler classifications (RandomForest). Remote Sensing of Environment, 100, 356−362.

Liaw, A., 2015. Package 'randomForest': Breiman and Cutler's Random Forests for Classification and Regression. https://cran.r-project.org/web/packages/randomForest/randomForest.pdf. Last accessed: 1 Jul 2017.

Masek, J., E. Vermote, N. Saleous, R. Wolfe, F. Hall, K. Huemmrich, F. Gao, J. Kutler, and T. Lim, 2006. A Landsat surface reflectance dataset for North America, 1990-2000 RID F- 3944–2010, IEEE Geoscience and Remote Sensing Letters, 3(1):68–72.

McFeeters, S., 1996. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. Int. J. Remote Sens. 17, 1425–1432.

Moore, R.T. & Hansen, M.C., 2011. Google Earth Engine: a new cloud- computing platform for global-scale earth observation data and analysis. American Geophysical Union, Fall Meeting, abstract #IN43C- 02.

Na, X., Zhang, S., Li, X., Yu, H., Liu, C., 2010. Improved land cover mapping using Random Forests combined with Landsat Thematic Mapper Imagery and ancillary geographic data. Photogramm. Eng. Remote Sens. 76 (7), 833–840

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

Pervez, M.S., Brown, J.F., 2010. Mapping irrigated lands at 250-m scale by merging MODIS data and National Agricultural Statistics. Remote Sens. 2, 2388–2412.

Plaza, a J., Tilton, J.C., 2005. Automated Selection of Results in Hierarchical\nSegmentations of Remotely Sensed\nHyperspectral Images. Process. Imagens/Classificação 4946–4949.

Portmann, F.T., Siebert, S., Döll, P., 2010. MIRCA2000—Global monthly irrigated and rainfed crop areas around the year 2000: A new high-resolution data set for agricultural and hydrological modeling. Global Biogeochem. Cycles 24, 1–24.
Pouliot, D., Latifovic, R., Zabcic, N., Guindon, L., Olthof, I., 2014. Development and assessment of a 250m spatial resolution MODIS annual land cover time series (2000-2011) for the forest region of Canada derived from change-based updating. Remote Sens. Environ. 140, 731–743.

Ramankutty, N., Evan, A.T., Monfreda, C., Foley, J.A., 2008. Farming the planet: 1. Geographic distribution of global agricultural lands in the year 2000. Global Biogeochem. Cycles 22, 1–19.

Rodriguez-galiano, V.F., Chica-olmo, M., Abarca-hernandez, F., Atkinson, P.M., Jeganathan, C., 2012. Remote Sensing of Environment Random Forest classi fi cation of Mediterranean land cover using multi-seasonal imagery and multi-seasonal texture. Remote Sens. Environ. 121, 93–107.

Rogan, J., Franklin, J., Roberts, D.A., 2002. A comparison of methods for monitoring multitemporal vegetation change using Thematic Mapper imagery. Remote Sens. Environ. 80, 143–156

Rouse, J.W.,Haas, R.H., Schell, J.A., Deering, D.W., 1973. Monitoring vegetation systems in the Great Plains with ERTS. Third ERTS Symp., Washington, NASA 309–317

Schneider, P., Roberts, D.A., Kyriakidis, P.C., 2008. A VARI-based relative greenness from MODIS data for computing the Fire Potential Index. Remote Sens. Environ. 112, 1151–1167.

Servicio de Información Agroalimentaria y Pesquera (SIAP), 2017. National Coverage of Agricultural Borders Series II. Mexico. https://datos.gob.mx/busca/dataset/frontera-agricola-serie-ii. Last accessed: 1 May 2017.

Simonetti, D., Simonetti, E., Szantoi, Z., Lupi, A., Eva, H.D., 2015. First results from the phenology-based synthesis classifier using Landsat 8 imagery. IEEE Geosci. Remote Sens. Lett. 12 (7), 1–5.

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

Szuster, B.W., Chen, Q., Borger, M., 2011. A comparison of classification techniques to support land cover and land use analysis in tropical coastal zones. Applied Geography 31, 525–532.

Teluguntla, P., Thenkabail, P.S., Xiong, J., Gumma, M.K., Giri, C., Milesi, C., Ozdogan, M., Congalton, R., Tilton, J., Sankey, T.R., Massey, R., Phalke, A., Yadav, K., 2015. 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. Chapter 6. In: Thenkabail, P.S. (Ed.), Remote Sensing Handbook. Volume II: Land Resources: Monitoring, Modeling, and Mapping. Taylor and Francis, New York, USA, pp. 131–159.

Thenkabail, P.S., 2010. Global croplands and their importance for water and food security in the twenty-first century: Towards an ever green revolution that combines a second green revolution with a blue revolution, Remote Sensing, 2(9):2305–2312
Thenkabail, P.S., Biradar, C.M., Noojipady, P., Dheeravath, V., Li, Y., Velpuri, M., Gumma, M., Gangalakunta, O.R.P., Turral, H., Cai, X., Vithanage, J., Schull, M. a., Dutta, R., 2009. Global irrigated area map (GIAM), derived from remote sensing, for the end of the last millennium. Int. J. Remote Sens. 30, 3679–3733.

Thenkabail, P.S., M.A. Hanjra, V. Dheeravath, and M. Gumma, 2011. Global croplands and their water use remote sensing and non-remote sensing perspectives, Advances in Environmental Remote Sensing: Sensors, Algorithms, and Applications (Q. Weng, editor), Taylor and Francis, Boca Raton, Florida, pp. 383–419.

Tilton, J. C., 2016. RHSeg User's Manual: Including HSWO, HSeg, HSegExtract, HSegReader, HSegViewer and HSegLearn, version 1.64, Mar. 22, 2016. Available on request: James.C.Tilton@nasa.gov

Tilton, J., Tarabalka, Y., Montesano, P., Gofman, E., 2012. Best Merge Region-Growing Segmentation with Integrated Nonadjacent Region Object Aggregation 50, 4454–4467.
Vapnik, V.N., 1995. The nature of statistical learning theory. Springer Science & Business Media.