



Mapping cropland extent of Southeast and Northeast Asia using multi-year time-series Landsat 30-m data using a random forest classifier on the Google Earth Engine Cloud



Adam J. Oliphant^{a,*}, Prasad S. Thenkabail^{a,*}, Pardhasaradhi Teluguntla^{a,b}, Jun Xiong^{a,b}, Murali Krishna Gumma^c, Russell G. Congalton^d, Kamini Yadav^d

^a Western Geographic Science Center (WGSC), U. S. Geological Survey (USGS), 2255, N. Gemini Drive, Flagstaff, AZ 86001, USA

^b Bay Area Environmental Research Institute (BAERI), NASA Research Park, Moffett Field, CA 94035, USA

^c RS/GIS Lab, Innovation Systems for the Drylands, International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, 502 324, India

^d University of New Hampshire, Dept. of Natural Resources & the Environment, Durham, NH, USA

ARTICLE INFO

Keywords:

Southeast Asia
Northeast Asia
Croplands
Food security
Water security
Landsat

ABSTRACT

Cropland extent maps are useful components for assessing food security. Ideally, such products are a useful addition to countrywide agricultural statistics since they are not politically biased and can be used to calculate cropland area for any spatial unit from an individual farm to various administrative units (e.g., state, county, district) within and across nations, which in turn can be used to estimate agricultural productivity as well as degree of disturbance on food security from natural disasters and political conflict. However, existing cropland extent maps over large areas (e.g., Country, region, continent, world) are derived from coarse resolution imagery (250 m to 1 km pixels) and have many limitations such as missing fragmented and/or small farms with mixed signatures from different crop types and/or farming practices that can be, confused with other land cover. As a result, the coarse resolution maps have limited usefulness in areas where fields are small (< 1 ha), such as in Southeast Asia. Furthermore, coarse resolution cropland maps have known uncertainties in both geo-precision of cropland location as well as accuracies of the product. To overcome these limitations, this research was conducted using multi-date, multi-year 30-m Landsat time-series data for 3 years chosen from 2013 to 2016 for all Southeast and Northeast Asian Countries (SNACs), which included 7 refined agro-ecological zones (RAEZ) and 12 countries (Indonesia, Thailand, Myanmar, Vietnam, Malaysia, Philippines, Cambodia, Japan, North Korea, Laos, South Korea, and Brunei). The 30-m (1 pixel = 0.09 ha) data from Landsat 8 Operational Land Imager (OLI) and Landsat 7 Enhanced Thematic Mapper (ETM+) were used in the study. Ten Landsat bands were used in the analysis (blue, green, red, NIR, SWIR1, SWIR2, Thermal, NDVI, NDWI, LSWI) along with additional layers of standard deviation of these 10 bands across 1 year, and global digital elevation model (GDEM)-derived slope and elevation bands. To reduce the impact of clouds, the Landsat imagery was time-composited over four time-periods (Period 1: January–April, Period 2: May–August, and Period 3: September–December) over 3-years. Period 4 was the standard deviation of all 10 bands taken over all images acquired during the 2015 calendar year. These four period composites, totaling 42 band data-cube, were generated for each of the 7 RAEZs. The reference training data (N = 7849) generated for the 7 RAEZ using sub-meter to 5-m very high spatial resolution imagery (VHRI) helped generate the knowledge-base to separate croplands from non-croplands. This knowledge-base was used to code and run a pixel-based random forest (RF) supervised machine learning algorithm on the Google Earth Engine (GEE) cloud computing environment to separate croplands from non-croplands. The resulting cropland extent products were evaluated using an independent reference validation dataset (N = 1750) in each of the 7 RAEZs as well as for the entire SNAC area. For the entire SNAC area, the overall accuracy was 88.1% with a producer's accuracy of 81.6% (errors of omissions = 18.4%) and user's accuracy of 76.7% (errors of commissions = 23.3%). For each of the 7 RAEZs overall accuracies varied from 83.2 to 96.4%. Cropland areas calculated for the 12 countries were compared with country areas reported by the United Nations Food and Agriculture Organization and other national cropland statistics resulting in an R^2 value of 0.93. The cropland

* Corresponding authors.

E-mail addresses: aoliphant@usgs.gov (A.J. Oliphant), pthenkabail@usgs.gov (P.S. Thenkabail).

<https://doi.org/10.1016/j.jag.2018.11.014>

Received 6 August 2018; Received in revised form 24 November 2018; Accepted 28 November 2018

Available online 22 May 2019

0303-2434/ Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

areas of provinces were compared with the province statistics that showed an $R^2 = 0.95$ for South Korea and $R^2 = 0.94$ for Thailand. The cropland products are made available on an interactive viewer at www.croplands.org and for download at National Aeronautics and Space Administration's (NASA) Land Processes Distributed Active Archive Center (LP DAAC): <https://lpdaac.usgs.gov/node/1281>.

1. Introduction

The spatial distribution of cropland areas is continually changing due to many factors such as climate variability resulting in intra- and inter-annual changes in temperature and precipitation, as well as demographic changes involving migration of rural farming communities to urban areas in many parts of the world, urbanization expansion into croplands, and geopolitical issues. Knowing the precise location of the cropland areas, the ability to map every farm, small or large, with sufficiently high resolution (30-m or better) over very large spatial extents with high degrees of accuracy is of great importance to study, access, and plan global food and water security scenarios in an increasingly inter-connected world (Teluguntla et al., 2016; Thenkabail et al., 2010; Waldner et al., 2016, 2015). Indeed, cropland area mapping with great geographic precision as well as accuracies in high resolution (30-m or better) is the first step in developing accurate understanding, modeling, and mapping of higher level cropland products such as crop types, cropping intensities, and watering methods (e.g., irrigated or rainfed). All of these products play a key role in modeling and mapping crop productivity and crop water-productivity, which are crucial in food and water security studies (Gerbens-Leenes and Nonhebel, 2004).

The need to produce a 30-m cropland extent product over Southeast Asia and Northeast Asia is of great importance (Asian Development Bank, 2009; FAO, 2016) and brings in unique challenges (Hurni et al., 2017; Li et al., 2017; Reid et al., 2013; Zhou et al., 2016). Southeast Asia has some of the most favorable climates in the world for agriculture, with large areas of Thailand, Cambodia, Vietnam, Philippines, and Indonesia able to support agriculture year-round. The main crop is rice and governments have put a great emphasis on national agricultural self-sufficiency, putting great effort in developing infrastructure and subsidies to support rice production (Panuju et al., 2013). Existing croplands not only feed the population of these countries, but also play a key role in economic income of the nations with many food products such as rice, coffee, tea, cocoa, and palm oil exported to many countries in the world. As populations and markets grow, so too does the demand for food. However, expansion of cropland area likely will be unfeasible due to the need for increase in land for urban development (Bren d'Amour et al., 2016; Bruinsma, 2009; Mutert and Fairhurst, 2002), wildlife habitat conservation (Li et al., 2016), ecosystem services, and submergence and sea level rise (Meybeck et al., 2012). As a result, these factors necessitate more complete and sustainable utilization of existing farmland (Thenkabail et al., 2010).

The landscape in Southeast Asia is a patchwork of different land uses and land covers, which presents difficulties when mapping at 250 m to 1 km Moderate-resolution Imaging Spectroradiometer (MODIS) resolutions (Suepa et al., 2016). The majority of existing cropland extent products (see Teluguntla et al., 2016) over very large areas such as continents or the entire world are coarse resolution (≥ 250 m) with significant uncertainties associated with the precise location of croplands as well as their accuracies. The majority of cropland maps derived by remote sensing for Southeast or Northeast Asia are limited to 250-m to 500-m ground resolution, about 6.25–25 hectares (ha) per pixel, because these maps rely on MODIS data (Ahamed and Bolten, 2017; Gumma et al., 2018; Setiawan et al., 2013; Sianturi et al., 2018; Tingting and Chuang, 2010) or other moderate resolution sensors (SPOT 1 km) (Manjunath et al., 2015). Further, they

fail to capture large proportions of individual small farms (errors or omissions) or they capture a significant proportion of non-croplands as croplands (errors of commissions).

There are a few existing high resolution (30-m or better) cropland products for small parts of Southeast and Northeast Asia (Cheng et al., 2016; Hurni et al., 2017; Kontgis et al., 2015; Sonobe et al., 2017; Torbick et al., 2016), but none for the entire area. Existing cropland studies often cover small areas ranging from 1000 km² (a small portion of a Landsat image) to a few Landsat footprints. Such small maps are highly inadequate for monitoring food security on a regional or national scale. For products to be useful for decision makers, they need to cover sufficiently large areas at high spatial resolution (30-m or better).

Current cropland classifications of Southeast Asia either are of insufficient spatial resolution or extent or of insufficient quality. A small number of global land use/land cover (LULC) products were created at moderate resolution, however, croplands are just one of the LULC classes without an attempt to capture croplands in particular, such as GLC2000 (Bartholomé and Belward, 2005), GlobCover (Arino et al., 2007), GLC-SHARE (Latham et al., 2014), and MODIS Land Cover (Friedl et al., 2002). In these large-scale land cover products, cropland is intermixed with pastures which limits their usefulness for deriving areas devoted to food production or as an agriculture mask for further work. However, high-quality and high-resolution cropland maps have been produced for continents outside Southeast Asia. Xiong et al. (2017b) produced the first large-scale Landsat-derived 30-m cropland extent product over the entire continent of Africa and Teluguntla et al. (2018) produced an excellent Landsat-derived 30-m cropland extent product for all of China and Australia.

In the last few decades, multi-temporal classification of satellite imagery has become an important tool in LULC science at regional, national, continental, and global scales (Chen et al., 2018; Gallego et al., 2014; Giri et al., 2003). Until recently, such analysis at continental and global scales were restricted to coarse resolution imagery like Advanced Very-High-Resolution Radiometer (AVHRR) 1-km and MODIS 250 m–500 m (Gumma et al., 2016; Teluguntla et al., 2017; Xiong et al., 2017a). Due to the expansion of parallel processing and huge cloud computing data centers, it is now possible to create global classified maps using Landsat 30-m imagery as demonstrated by studies on Global Forest Cover, (Hansen et al., 2013) and GlobeLand30 (Chen et al., 2015). Also, a major limitation with mapping with Landsat imagery is cloud cover; some of the cloudiest regions on Earth are in Southeast Asia (Kontgis et al., 2015; Li et al., 2017). This cloudiness necessitates using advanced cloud identification and compositing algorithms to overcome errors associated with clouds and cloud shadows (Xiong et al., 2017b). Availability of the Google Earth Engine (GEE) cloud computing platform allows processing of massively large volumes of multi-temporal satellite data from high spatial resolution imagery such as the Landsat and Sentinel sensors (Gorelick et al., 2017). GEE has been used for many global and continent-wide land cover analyses with great success (Dong and Xiao, 2016). The chief benefit of using GEE is it alleviates the burden of downloading and processing the raw imagery locally; instead processing is done on the cloud (Gorelick et al., 2017).

Given the above background, the overarching goal of this research was to produce a precise and accurate cropland extent product of Southeast and Northeast Asia using Landsat 30-m data, within the GEE cloud computing platform, using machine learning algorithms (MLA) (Teluguntla et al., 2018; Xiong et al., 2017b). This study helps

determine consistent and objective cropland areas at the national and sub-national level and compare the same with the national and-sub-national statistics obtained from the country statistics.

2. Study area

2.1. Study area location

This study covers a total of 12 countries: 9 Association of Southeast Asian Nations (ASEAN) countries (Brunei, Cambodia, Indonesia, Laos, Malaysia, Myanmar, Philippines, Thailand, and Vietnam), and 3 Northeast Asian countries (Japan, North Korea, and South Korea) (Fig. 1). Singapore was not included since the total agricultural area is very small (about 1000 ha) and very technologically advanced, making it atypical of agriculture in the vast majority of Southeast Asia (Agri-Food and Veterinary Authority of Singapore, 2018). Together, we refer to the 12 countries as Southeast and Northeast Asian countries or SNAC (Fig. 1). There are 836.7 million people in SNAC, which is about 11.3% of the world's 7.6 billion people (UN DESA 2015).

2.2. Study area stratification and study area characteristics

A major challenge with mapping over such large areas is that the agricultural practices vary greatly from region to region (Brown et al., 2012; Portmann et al., 2010). Additionally, seasonal cloud cover during monsoons varies between regions and locales. Since these areas cover vast stretches of complex landscapes with varying climate, topography, and geography, we decided to divide SNAC into seven refined agro-ecological zones (RAEZs) (Table 1). The seven RAEZs were created taking the following into consideration. Initially, we started with the Köppen Climate map (Peel et al., 2007) and UN Food and Agriculture Organization (FAO) agro-ecological zones (AEZs) (Fischer et al., 2000). However, much of the area fell in tropical and subtropical regimes,

Köppen classes Af and Am (Af: Equatorial climate, Am: Monsoon climate). There was a lack of sufficient distinction within these zones to account for the distinct agricultural patterns and variations in climate, topography, vegetation, soils, and slope, which are all important factors in cropland extent classifications. To overcome this lack of distinction, we reviewed additional spatial data layers like soils, elevation, slope, and political administrative boundaries along with AEZs to arrive at seven broad RAEZs (Table 1), using methodology similar to Xiong et al. (2017b). Attributes of these seven RAEZs for SNAC follow.

3. Methods

3.1. Definition of cropland extent

Any mapping should begin with a clear definition on what is being mapped. When definitions vary between different mapping projects, they cause one of the greatest uncertainties in the inter-comparison of the products. So, when two maps are compared, an understanding of the differences in definitions is key. For this project on Global Food Security-support Analysis Data project at 30-m for the 12 countries of Southeast and Northeast Asia (GFSAD30SEA), cropland extent map was defined as: “lands cultivated with plants harvested for food, feed, and fiber, including both seasonal crops (e.g., wheat, rice, corn, soybeans, cotton) and continuous plantations (e.g., coffee, tea, rubber, cocoa, oil palms)” (Teluguntla et al., 2016). Additionally, since the goal of this work was to create a cropland extent map that could be used for multiple years, cropland fallows were included in the cropland extent class to account for land left fallow as a result of drought conditions, crop rotations, and other causes (e.g., price drop for some crop produce in a given year). Farmlands that are not cultivated during a season or a year are considered cropland fallows. Simply put, this cropland extent map includes standing annual crops + permanent plantation crops + cropland fallows. A major benefit of using Landsat is its spatial resolution of 30 m

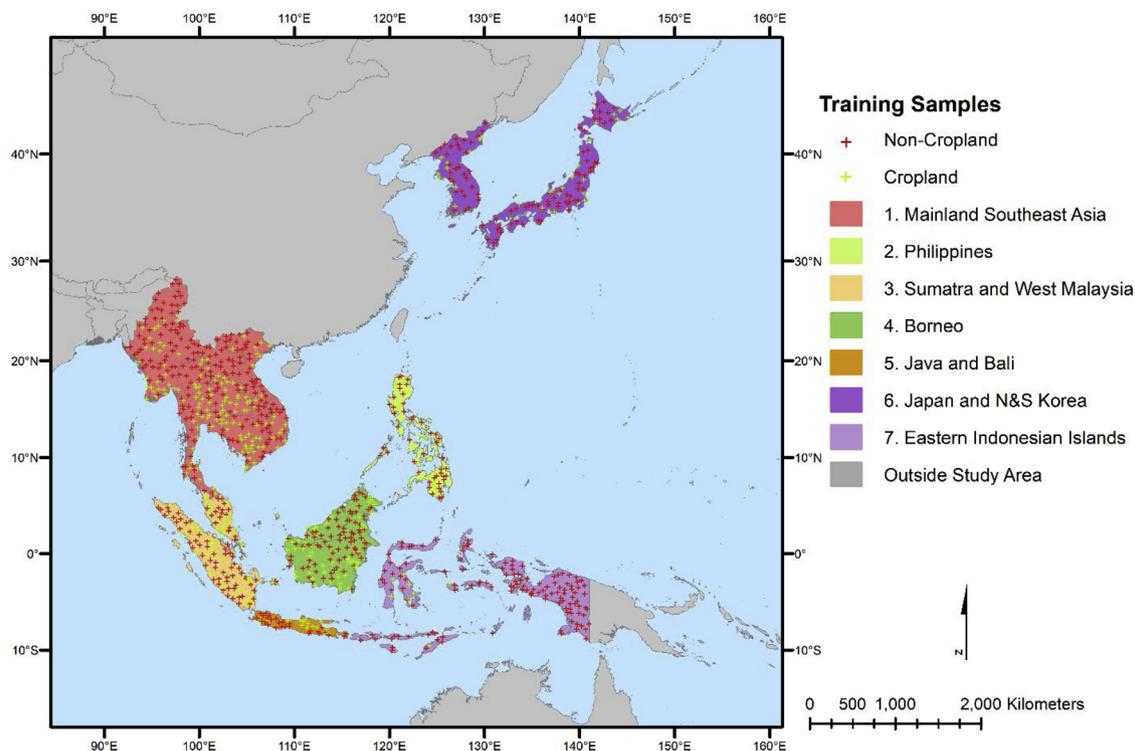


Fig. 1. Southeast and Northeast Asian countries (SNACs) were divided into 7 refined agro-ecological zones (RAEZs) for the pixel-based supervised random forest cropland extent classification. RAEZs were chosen based on temperature, political boundaries, seasonal precipitation, and farming practices. The zones are as follows: RAEZ 1 - Mainland SE Asia; RAEZ 2 - Philippines; RAEZ 3 - Sumatra and West Malaysia; RAEZ 4 - Java and Bali; RAEZ 5 - East Malaysia, Kalimantan, and Brunei; RAEZ 6 - Northeast Asia; RAEZ 7 - Eastern Indonesian Islands. Also shown are reference samples of cropland in green and non-cropland in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

Table 1

Table with Revised Agro-Ecological Zones (RAEZs) #, RAEZ name, Köppen climate classification, temperature mean annual, average temperature of coldest and warmest month, precipitation mean annual, average precipitation of coldest and warmest month. Average temperature and precipitation are over years 1991–2015, sourced from [World Bank, 2016 Climate Change Knowledge Portal](#)***.

RAEZ #	RAEZ Name	Dominant Köppen Class	Average temp. of coldest month °C	Average temp. of warmest month °C	Mean average annual temp. °C	Average precip. of annual rainfall mm	Average precip. of wettest month mm	Average precip. of driest month mm	Average monthly precip. mm
1	Mainland SE Asia	Af, Am, Csa	21.5	27.3	25.2	1739	304	19	145
2	Philippines	Af, Am	24.4	27	25.9	2536	310	101	211
3	Sumatra & W. Malaysia	Af	25.2	26.3	25.8	3105	340	190	259
4	Java & Bali	Af, Am	25.9	26.5	26.2	2869	307	165	239
5	Borneo	Af	25.2	26.1	25.7	3296	337	198	275
6	Japan, N&S Korea	Cfa, Dfa, Dwa	−4.7	22.9	9.7	1354	265	37	113
7	Eastern Indonesian Islands	Af	24.3	26.5	25.4	1519	261	11	127

Af = Equatorial climate Am = Monsoon climate Aw = Tropical savanna climate.

Cfa = Humid subtropical climate Csa = Warm Mediterranean climate.

Dfa = Humid continental climate Dwa = Humid continental climate (dry winter).

*** Note: See [Fig. 1](#) for spatial coverage of RAEZ's.

(0.09 ha per pixel). At this resolution an overwhelming proportion of the world's croplands are higher resolution than 0.09 ha. However, even when cropland falls below this resolution, as is the case in some fragmented farms or in parts of the economically developing world, the temporal signal (e.g., reflectivity, NDVI) of various magnitude (e.g., depending on: crop types, proportion of the pixel occupying the crop, irrigated or rainfed) from sub-pixel fraction of farming within a Landsat pixel area of 0.09 ha, is still captured as croplands. In area calculations they form a full pixel area (FPA) of 0.09 ha. Accounting for sub-pixel fraction of these pixels is not required as such pixels form < 3% of the global cropland area ([Monfreda et al., 2008](#); [Thenkabail et al., 2007](#)).

3.2. Input imagery: landsat 30-m time-series data

The Landsat 8 and 7, 30-m, time-series data ([Table 2](#)) during the 2013–2016 time-period were used for mapping the cropland extent of SNAC. Landsat 8 and 7 satellites have high spectral calibrations and produce high quality and consistent data suitable for spatial analysis over very large areas. The Landsat satellite series are launched by the National Aeronautics and Space Administration (NASA) and the U.S. Geological Survey (USGS) freely processes and distributes the imagery to the public for no direct cost to users.

It was necessary to use both Landsat 8 and 7 imagery to maximize data coverage and to derive cloud-free pixels for analysis. Data are acquired over the study area every 8 days from the two satellites. Overall, 10 bands of data ([Table 2](#)) were used. The blue, green, red, near infrared (NIR), shortwave infrared (SWIR) 1, SWIR 2, and thermal bands along with the vegetation indices Normalized Difference Vegetation Index (NDVI), Normalized Burn Index (NBR2), and Liquid

Surface Water Index (LSWI) were used for this classification ([Table 2](#)). NDVI was selected to help distinguish dense vegetation including forests. NBR2 was chosen to distinguish barren and urban lands from other land cover. LSWI was included to help separate rice paddy and other bodies from land ([Kontgis et al., 2015](#)).

Four temporal periods ([Fig. 2](#)) were chosen for image composition based on regional cropping calendars and cloud cover. The periods chosen were: period 1 (day of year (DOY) 1 to 120); period 2 (DOY 121–240); period 3 (DOY 241 to 365). Period 4 was the standard deviation of all cloud-free pixels during the entire year, 2015, for all 10 bands. The standard deviation across 1 year was chosen to capture inter-annual variability and to focus the classification on one year (2015). The composites were compared to sub-meter to 5-m very high spatial resolution imagery (VHRI) to ensure that no artifacts were introduced into the composites (such as cloud cover). This comparison was done to ensure that no erroneous classification results due to missing data occurred in portions of the image ([Teluguntla et al., 2018](#)). Through experimentation, we found to ensure that gapless cloud-free images could be generated for the entire SNAC area ([Fig. 1](#)) for each period, it was necessary to generate composites over multiple years. Although having composites within only a year would be ideal, analysis of temperature and precipitation data from the Climate Change Knowledge Portal ([World Bank, 2016](#)) indicated that the years 2013 to 2016 contained dry, average, and wet years ([World Bank, 2016](#)). This was of benefit in that it encapsulated the inter-annual variability in crop extent due to variations in precipitation and rainfall. Additionally, no areas had image gaps in agricultural areas when using 3-year composites, which was not the case when using 2 years of data, particularly in very cloudy areas found in Borneo and Sumatra.

Table 2

Characteristics of Multi-temporal Landsat 7 and 8 data used in the study.

Band Name	Landsat 8 OLI Spectral Range μm	Landsat 7 ETM + Spectral Range μm	Vegetation Index (VI) Name	Equation
Blue	0.452 – 0.512	0.45 – 0.52	NDVI	$(\text{NIR} - \text{Red}) / (\text{NIR} + \text{Red})$
Green	0.533 – 0.590	0.52 – 0.60		
Red	0.636 – 0.673	0.63 – 0.69	NBR2	$(\text{SWIR1} - \text{SWIR2}) / (\text{SWIR1} + \text{SWIR2})$
NIR	0.85 – 0.879	0.77 – 0.90		
SWIR 1	1.566 – 1.651	1.55 – 1.75	LSWI	$(\text{NIR} - \text{SWIR1}) / (\text{NIR} + \text{SWIR1})$
SWIR 2	2.107 – 2.294	2.09 – 2.35		
Thermal	10.60 – 11.19	10.40 – 12.50		

Note: NIR = near infrared, SWIR = shortwave infrared, OLI = Operational Land Imager.

ETM+ = Enhanced Thematic Mapper plus, NDVI = normalized difference vegetation index.

NBR = normalized burn ratio, LSWI = land surface water index.

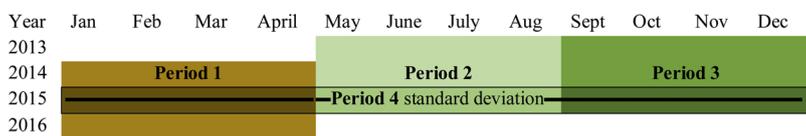


Fig. 2. Time period of Landsat 8 and 7 imagery used in the study. Landsat 7 and 8 images acquired from May 1, 2013, through April 30, 2016, were used to composite and mosaic nominal 30-m data used in cropland classification. Note: Period 4 was the standard deviation calculated over all images acquired over 2015 for each band.

In addition to 40 Landsat 7 and 8 derived bands of data [10 bands (Table 2) x 4 periods (Fig.2)], slope and elevation were used in this classification because croplands vary based on elevation and slope (Iizumi and Ramankutty, 2015; Reda and Tripathi, 2016). The September 2014 release of the nominal 30-m resolution Shuttle Radar Topography Mission (SRTM) Version 3 void filled global elevation data was used to determine slope and elevation (Farr et al., 2007; Kobrick and Crippen, 2014).

The methodology used to create the imagery stack or mega file data cube (MFDC) is described below and illustrated graphically in Fig. 3. All processing was performed on the Google Earth Engine (GEE) cloud computing platform to enable seamless and fast computing and the ability to handle massive amounts of data over very large areas. A custom cloud detecting script available at (Oliphant et al., 2017a) was run within GEE to mask out cloud-impacted pixels. Top of the Atmosphere (TOA) images were used instead of Surface Reflectance (SR) images due to low SR imagery availability in GEE. The novel script ran faster and masked more cloud impacted pixels than Fmask (Housman et al., 2015; Zhu et al., 2015). These 42 bands (Fig. 3) were generated separately for each of the 7 RAEZs and formed the baseline remote sensing data used in the cropland versus non-cropland classification in RAEZ.

3.3. Reference training data

Reference data were gathered from multiple sources: (1) Sub-meter to 5-m VHRI acquired circa 2013 to 2016 and made available to us from the National Geospatial Agency (NGA), (2) Ground data through extensive field visits, and (3) Tertiary published data from collaborators. The cropland versus non-cropland reference training and validation data used to classify imagery using the random forest (RF) machine learning algorithm (MLA) were gathered from VHRI such as WorldView, QuickBird, and GeoEye. We generated the reference training data for RF MLA by initially creating at least 500 random samples for each of the 7 RAEZs. Each sample was extracted from a 90 m x 90 m (Fig. 4) area to ensure the validity of the sample in terms of its homogeneity and geographic integrity. Additional training samples were added to the initial selection of 500 random samples in areas where it was obvious that the classified images did not visually match the landcover observed from multiyear VHRI (explained in detail in section 3.6). Ultimately, there was a total of 7849 training samples (Table 3) from VHRI spread across SNAC. Their distribution in each of the 7 RAEZs is summarized in Table 3. The entire sample training and validation dataset used in this study is available at the following website: <https://croplands.org/app/>

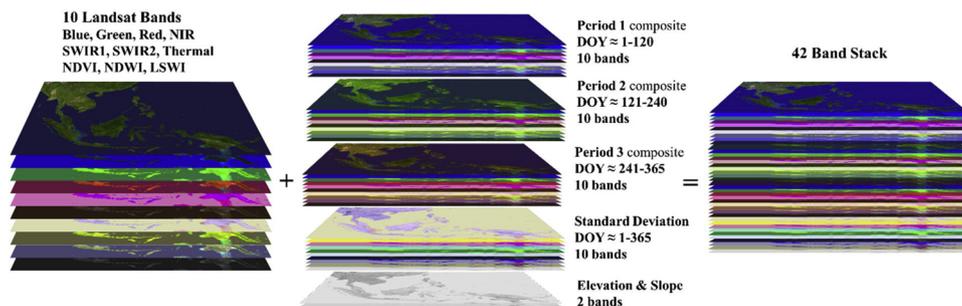


Fig. 3. Data cube of Landsat 8 and 7 for each of the three periods, taking 2013–2016 imagery over entire Southeast and Northeast Asian countries. A 10-band (blue, green red, NIR, SWIR1, SWIR2, Thermal, NDVI, NDWI, LSWI) median composite of Landsat 8 and 7 was created for four time periods (1: January–April, 2: May–August, 3: September–December). In addition, 10 standard deviation bands from each image collected in 2015 (period 4) and slope and elevation derived from SRTM were also used. This resulted in a 42-band data stack for each of the seven zones in the study area which

were inputs for the random forest crop extent classifications (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

data/search.

In addition, over 1960 ground reference samples were collected through extensive field campaigns in Thailand, Myanmar, Vietnam, and Indonesia by the authors (Gumma et al., 2018). In particular, our field visit to Thailand and Indonesia in 2016 greatly informed our understanding in regard to crop rotations and the heterogeneity of small farms (e.g. farmers commonly have a small, deep pond they use for irrigation water storage and aquaculture). The standard process of gathering field data is extensively described in Teluguntla et al. (2017). The same process was followed here. Further, tertiary collaborators provided us with 2383 tertiary reference samples in Indonesia, Japan, South Korea, Thailand, and Vietnam (Bui et al., 2013; Dutta et al., 2015; Seong et al., 2008; Sharma et al., 2016). The ground data and the tertiary data were used in class identification and labeling. These data helped us improve the RF classification performance leading to optimal results.

3.4. Knowledge generation for random forest (RF) classifier

After reference training data were collected (Table 3), they were used to generate croplands versus non-croplands knowledge using the 42-band data cube (Fig. 3). These data are plotted in a bar and whiskers plot as illustrated for RAEZ 2 (Fig. 5). This knowledge was used in the RF algorithm to classify images. These knowledge plots provide clear indications in which bands there is separability between cropland versus non-croplands. Two box and whiskers knowledge plots are illustrated in Fig. 5 for RAEZ 2, Philippines. Cropland is shown in green (the first box) and non-cropland is shown in gold (the second box). The central line is the mean of all the values, while the top and bottom of the box are the 25th and 75th percentiles, respectively. The ends of the whiskers are defaults as defined in (Wickham and Chang, 2016), and the black dots are individual outlier samples. In order for all the values to fit legibility into the same plot, the following scaling was performed for visualization: temperature was converted from K to °C then divided by 25; slope was divided by 45 to convert percent slope to percent rise; elevation was divided by 1000.

3.5. Cloud computing on Google Earth Engine (GEE)

Due to the enormous size of the Landsat 30-m time-series data over very large areas such as SNAC (Fig. 1), it is essential to have a powerful platform for image organization, processing, and classification. In this study, we performed this task on the GEE cloud computing platform. GEE has the entire Landsat archive along with many publicly available

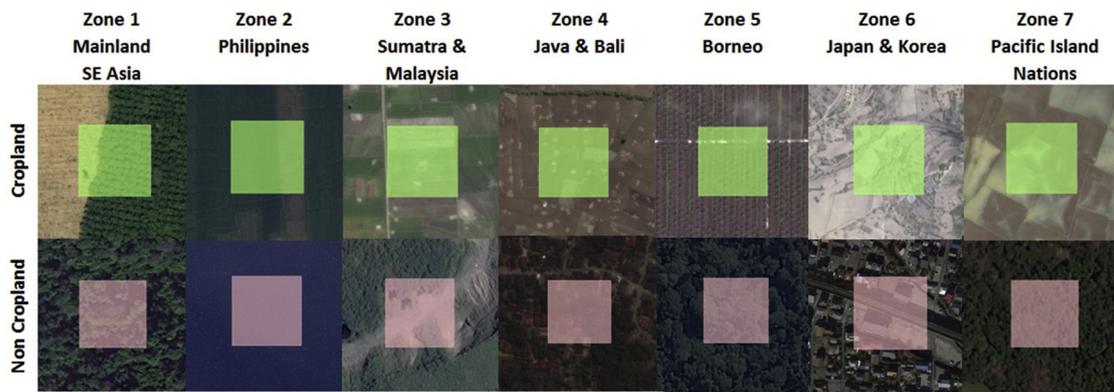


Fig. 4. Illustration of the reference training data collected from sub-meter to 5-m very high spatial resolution imagery (VHRI). The sub-meter to 5-meter VHRI sourced training samples for croplands *versus* non-cropland for use in the random forest algorithm in each of the 7 refined agro-ecological zones. Green and pink squares are 90 m × 90 m (9 Landsat pixels) for croplands and non-croplands respectively. The background is a VHRI from sensors such as WorldView, QuickBird, and GeoEye that shows a 250 m × 250 m area of ground (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

raster datasets from NASA, European Space Agency (ESA), and other imagery. GEE allows code to be brought to data; complex multi-temporal continental-scale data can be run using relatively simple JavaScript or Python code that can be shared and replicated by other researchers, lowering the barriers to utilizing supercomputers to perform geospatial analysis (Gorelick et al., 2017). The entire classification process was performed in GEE. First, code was executed that created cloud-free image composites. After the composites were generated, median pixel values corresponding to the training data extracted in GEE were used as the basis for the pixel-based supervised RF MLA classification. The results were viewed within GEE, and performance of classification was tested using reference data (e.g., field data from ground, data from collaborators, and VHRI; section 3.3), and this process was iterated using additional training data until optimal classification results of croplands *versus* non-croplands were attained. Post processing was performed within GEE and cropland areas were calculated in GEE based on the Global Administrative Unit Layers (GAUL) database by the United Nation’s (UN) Food and Agricultural Organization (FAO) (http://www.foodsec.org/tools_gaul.htm) country political boundaries.

3.6. Random forest algorithm

A pixel-based supervised random forest (RF) machine learning algorithm (MLA) was used for classification in the GEE cloud computing platform. Random forest classifiers applied to Landsat imagery in GEE have successfully mapped a variety of vegetation in SNAC including plantations such as oil palms (Lee et al., 2016), and rubber (Beckschäfer, 2017). The RF classifier is a relatively fast, nonlinear classifier that excels in producing good results from noisy data (Pelletier et al., 2016; Rodriguez-Galiano et al., 2012). It uses multiple decision

trees to assign classification labels. To reduce overfitting, each tree only classifies a subsection of training data. For a detailed description of the RF classifier, the reader is referred to (Breiman, 2001; Breiman and Cutler, 2004; Oliphant et al., 2017b). Through experimentation, we determined that 300 trees was a good balance between classification speed and accuracy. The default values were chosen for variablesPerSplit ($\sqrt{n_bands}$), minLeafPopulation (1), and bag Fraction (0.5). The outOfBagMode = true was used in order to use different random subsamples of training data in generating trees to reduce model overfitting.

As with most pixel-based supervised classifiers, obtaining quality results requires high quality (cloud free) input rasters (Fig. 3) and adequate training data (section 3.3, Table 3) that encompass the variability of the landscape. When such large regions are mapped, a large number of samples (Table 3) is required. An iterative approach was used to add and remove training samples to improve the map classification (Fig. 6). Each map was visually assessed to see how well the classified map correlated with observed cropland in the landscape from sub-meter to 5-m satellite VHRI, (Oliphant et al., 2014) primarily from WorldView 3. The number of iterations required to achieve satisfactory classification was related to the complexity of the area. The RF MLA was used to do a binary classification in each of the 7 RAEZs to generate cropland and non-cropland classes (Fig. 6).

3.7. Classification post processing

Post-processing image enhancements were used to create the final cropland extent product. In pixel-based land cover classification projects, large homogenous areas of one land cover class can often contain single pixels or small numbers of pixels of another land cover class.

Table 3

Number of reference training samples of croplands *versus* non-croplands in each of the 7 refined agro-ecological zones (RAEZs) in Southeast and Northeast Asian countries.

RAEZ#	RAEZ Name	Cropland Training Samples #	Non Cropland Training Samples #	Total Training Samples #	Land Area km ²	Area per Sample km ² /#
1	Mainland SE Asia	1,326	1,267	2,593	1,939,900	748
2	Philippines	350	330	680	300,000	441
3	Sumatra & W. Malaysia	305	317	622	604,100	971
4	Java & Bali	298	272	570	134,100	235
5	Kalimantan	257	349	606	743,300	1227
6	Japan and N&S Korea	460	481	941	598,700	636
7	Eastern Indonesian Islands	614	1,223	1,837	750,800	409
	Total	3,610	4,239	7849	5,070,900	646

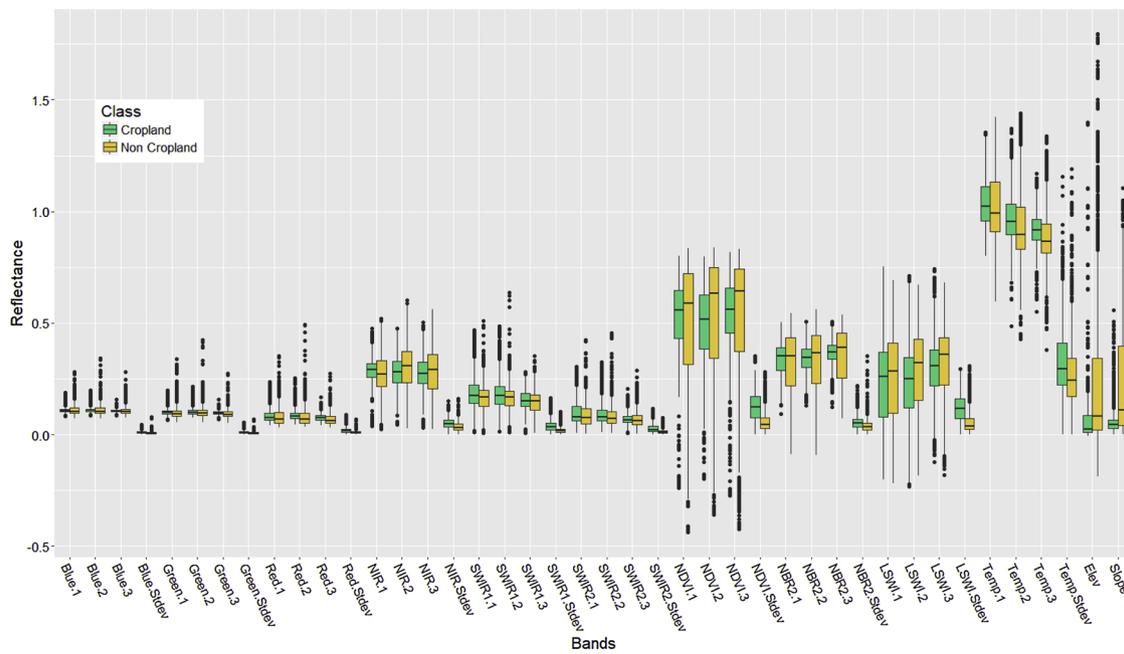


Fig. 5. Illustration of knowledge generation for croplands versus non-croplands for the refined agro-ecological zone 2 (RAEZ 2), Philippines. Knowledge generation for croplands (left in green) versus non-croplands (right bar in gold) developed based on reference training samples using the 42-band data cube. This knowledge is used in the random forest algorithm to classify cropland versus non-croplands. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

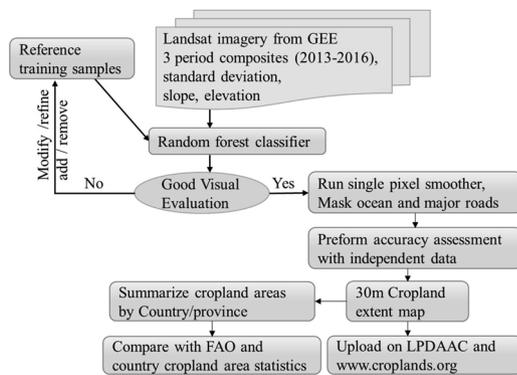


Fig. 6. Flowchart for mapping cropland extent in Southeast and Northeast Asia using pixel-based random forest machine learning algorithm.

Algorithms can be applied to remove these erroneous artifacts often called ‘Salt and Pepper’ noise, thus improving the accuracy and the visual correctness of map products. A disadvantage of using post classification smoothers is that fine features such as roads are obscured in the smoothed product although they were present in the pre-smoothed

Table 4

Reference validation data. Number of validation samples of croplands for versus non-croplands in each of the 7 refined agro-ecological zones (RAEZs) in Southeast and Northeast Asian countries.

RAEZ#	RAEZ Name	Cropland Validation Samples #	Non Cropland Validation Samples #	Total Validation Samples #	Percent of cropland for total area %
1	Mainland SE Asia	96	154	250	47.6
2	Philippines	56	194	250	7.2
3	Sumatra & W. Malaysia	67	183	250	14.5
4	Java & Bali	102	148	250	4.7
5	Borneo	71	179	250	14.4
6	Japan and N&S Korea	57	193	250	6.7
7	Eastern Indonesian Islands	12	238	250	4.9
	Total	461	1,289	1750	100.0

product. To correct for this, major roads were masked out using Open Street Map Vector files (March 30, 2017) downloaded from <https://extract.bbbike.org>. Additionally, coastlines and large water bodies were given a separate classification to separate them from the non-cropland class using the GlobCover version 2.3 product (Bicheron et al., 2011).

3.8. Uniqueness of this study relative to previous studies

There were some similarities, but also important differences between in this study relative to other recent studies (Teluguntla et al., 2018; Xiong et al., 2017b). Similarities/overlap in the methods used here compared to those of Xiong et al. (2017b) and Teluguntla et al. (2018) include:

- 1 Use of a Random Forest supervised machine learning algorithm;
- 2 Use of Google Earth Engine cloud platform; and
- 3 Methods involved in validation data collection and accuracy assessment.

Uniqueness of this study relative to that of Xiong et al. (2017b) and Teluguntla et al. (2018) include:

- 1 Use of standard deviation bands (Section 3.4) in the random forest machine learning algorithm, which enhanced the performance of the random forest significantly.
- 2 Use of extensive field data from the trip to Thailand and Indonesia (see last paragraph in Section 3.3) where substantial effort was taken to collect training and validation data. This effort enhanced the quality of data fed to train the random forest machine learning algorithms, by making them more robust. These data also bring greater clarity and confidence in the validation data; and
- 3 Use of post-processing image enhancement codes (Section 3.7) that further improved the final product.

3.9. Accuracy assessment

The accuracy of a mapped product must take into account the precision of the input raster data on which the classification was generated as well as the locational precision of the validation data used to assess the product. In order to account for these uncertainties, a minimum mapping unit of 0.9 ha (nine 30-m x 30-m pixels) was chosen. An ideal way to assess such accuracies is through a balanced sampling approach by a team using data that were not available to the analysts (Congalton et al., 2017). This ensures the accuracy is not artificially inflated as a result of the training data being from the same source as the validation data and it is impossible to fit the classification to the validation data. For this study, 250 randomly distributed sampling locations were generated for each of the 7 RAEZs, completely independent of the producers of the SNAC cropland product. Each of these validation sampling locations was either uniform cropland or non-cropland over a 90 m x 90 m area. If the area in which the sample unit was located contained a mix of cropland and non-cropland, the sampling unit was moved to the nearest area of cropland. The samples were visually interpreted by two independent analysts and only the samples agreed on by both were retained. The high-resolution imagery

referenced was collected between 2013–2016. Table 4 shows the distribution of the cropland and non-cropland validation samples. There were 250 samples generated for each of the seven RAEZs, (1750 in total) so each RAEZ has its own accuracy error matrix, as well as an overall accuracy matrix for the continent overall accuracy; user's accuracy (or errors of commissions) and producer's accuracy (or errors of omissions) were calculated to assess the accuracy of the classified maps after application of the post-processing steps defined above (Congalton and Kass, 2008).

3.10. Area assessment

Food security studies require cropland areas to be generated at national and sub-national scales. This product, mapping croplands at 30-m resolution (1 pixel = 0.09 ha), is ideal for cropland area assessments at national and sub-national levels and even at village and individual farm levels. Cropland areas calculated from this study were compared with several other studies [FAO Cultivated Area (FAO, 2013), Monthly Irrigated and Rainfed Cropland Areas year 2000 (MIRCA2000) (Portmann et al., 2010), Global Rain-fed, Irrigated, & Paddy Croplands (GRIPC) (Salmon et al., 2015), Global Irrigated Area Map (GIAM) (Thenkabail et al., 2009) and Global Map of Rainfed Cropland Areas (GMCA) (Biradar et al., 2009)] based on availability of data for comparisons. The areas calculated for the study countries were contrasted with areas obtained from FAO. Sub-national statistics were obtained for the provinces of South Korea and Thailand and were compared with the area statistics generated by this study.

4. Results

This section summarizes the 30-m cropland extent product. First, the cropland product is displayed in totality, with select areas highlighted to show detail. Second, an assessment of accuracies of the 30-m

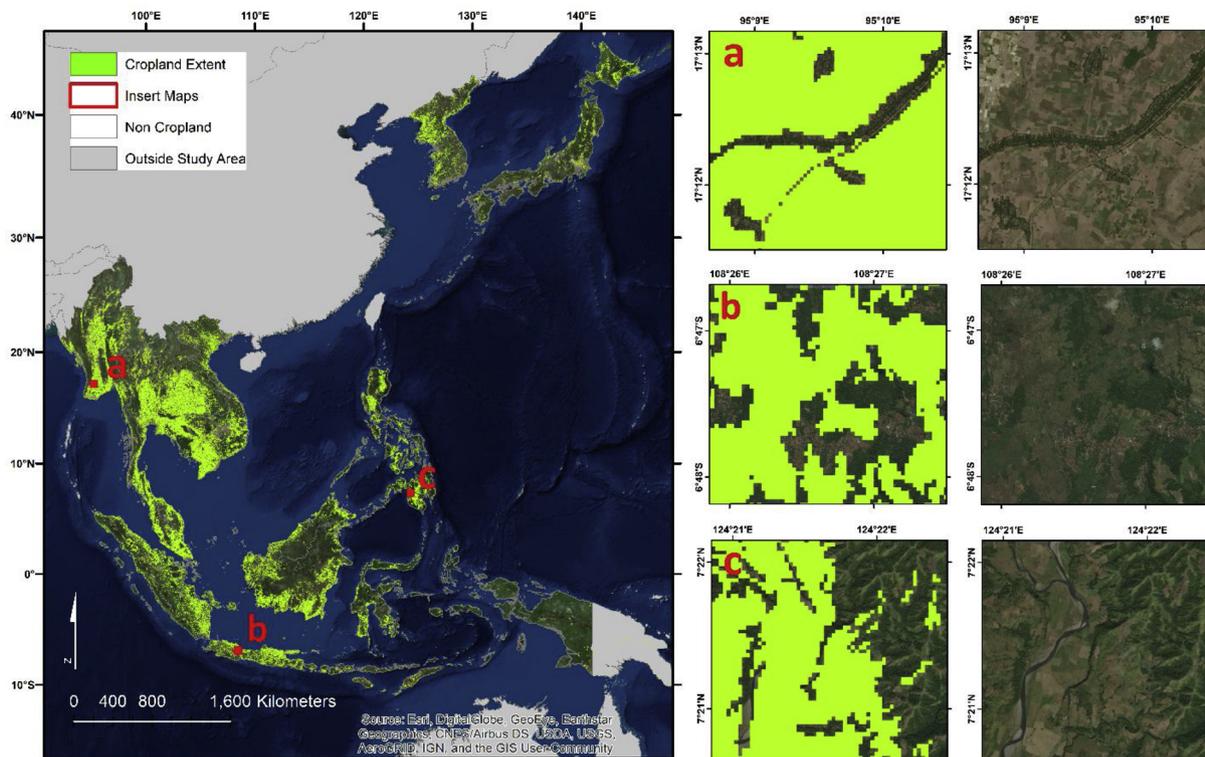


Fig. 7. The 30-m Landsat derived cropland extent product for the nominal year 2015 for Southeast and Northeast Asian countries. Green areas represent croplands and transparent areas are non-croplands. The product is viewable at full resolution at: www.croplands.org. This product, named GFSAD30SEACE, can be downloaded from: <https://lpdaac.usgs.gov/node/1279>. “Zoom-in” view at three locations: (a) Ayeyarwady, Myanmar, (b) Majalengka Regency, West Java, Indonesia, and (c) Maguindanao, Philippines. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

cropland extent product is presented. Third, cropland areas are computed for national and sub-national levels and are compared with the corresponding areas obtained from the national statistics.

4.1. 30-m cropland extent product

The study created a 30-m cropland extent product of the 12 countries (Fig. 7): nine Southeast Asian countries and three Northeast Asian countries for the nominal year 2015. The product was generated from a supervised random forest machine learning algorithm (Section 3.6) utilizing training data (Section 3.3), on seasonal custom cloud free image composites derived from Landsat 8 and 7 imagery (Section 3.2). It maps croplands at a very fine (1 pixel = 0.09 ha) spatial resolution over a very large area covering the entire SNAC in great detail (Fig. 7). The product is called the Global Food Security-support Analysis Data @ 30-m Cropland Extent for Southeast and Northeast Asia (GFSAD30SEA), and can be viewed at www.croplands.org and downloaded from the NASA LP DAAC: <https://lpdaac.usgs.gov/node/1279>.

An overview of the cropland extent of the entire SNAC is shown in Fig. 7 where cropland extent is shown in bright green and non-cropland extent is shown as transparent (no color). “Zoom-in” insert maps are shown for 3 areas labeled a, b, and c for small portions of Myanmar, Java, and Philippines respectively, to show map detail. These areas are magnified and shown along with VHRI for visual comparison.

4.2. Accuracies of the 30-m cropland extent product

The results showed that for the entire SNAC study area the overall accuracies were 88.6% with cropland class producer's accuracy of 81.6% (errors of omissions = 18.4%) and user's accuracies of 76.7% (errors of commissions = 23.3%) (Table 5). The column total lists the number of cropland and non-cropland samples that were generated by an independent validation team that were used to assess the accuracy of the product. For each of the 7 RAEZs overall accuracies varied between 83.2% and 96.4% with cropland class producer's accuracy of 67.2%–97.2% (errors of omissions = 2.8%–32.8%) (Table 6). Results clearly indicate the ability of Landsat data and methods used in this study to rapidly and accurately map croplands.

4.3. Total Net Cropland Areas (TNCAs)

The 30-m cropland extent product is ideal for generating cropland area statistics at the national as well as various sub-national levels. Table 7 shows cropland areas of the 12 countries compared with the cropland areas derived from a number of other sources including: MIRCA (Portmann et al., 2010); FAO (FAO, 2013); GRIPC (Salmon et al., 2015); and the combination of GIAM (Thenkabail et al., 2009) and GMRCA (Biradar et al., 2009).

The total net cropland area (TNCA) of 12 countries was 126.6 million hectares (Mha) as compared with MIRCA (132.2 Mha) and FAO national statistics (116.2 Mha). The cropland area calculated by GFSAD30 was 5.6 Mha less than MIRCA (4.2% difference) and 10.4 Mha greater than FAO (9.2% difference). This same information is shown graphically in Fig. 8 where GFSAD30 is compared to national agricultural areas reported by FAO.

4.4. Sub-national cropland areas

The 30-m resolution (1 pixel = 0.09 ha) allows the calculation of sub-national statistics down to a single farm level. This is indeed a great advantage of the 30-m product wherein one can generate statistics at various levels with very high degrees of accuracy. So, wherever national statistics were available at the sub-national level, we compared the GFSAD30SEACE (this study) 30-m derived statistics of sub-national level with sub-national level statistics available from the national ministries. Provincial cropland areas were obtained for South Korea

(Statistics Korea, 2017) and Thailand (National Statistical Office, 2013) and were compared with GFSAD30SEACE derived 30-m statistics, which are shown in Fig. 9 and Fig. 10 respectively. The large number of provinces in Thailand prohibited all provinces from being labeled in Fig. 10.

5. Discussion

With Southeast and Northeast Asia's population representing 11% of the world's population, and the number of people expected to increase from 837 million in 2017 to 958 million by 2050 (UN DESA, 2017), the complete utilization of existing farmland is of the utmost importance. Cropland extent maps are one way to demonstrate the current utilization of cropland (See et al., 2015; USDA NASS, 2014). In 2005 FAO estimated that 4% of the population of Southeast Asia suffered from severe food insecurity while an additional 14% suffered from moderate food insecurity (FAO, 2016). The situation is worse for stunted growth in children; in 2010 30% of the population under 5 years old were estimated to have stunted growth in Cambodia, Indonesia, Laos, Myanmar, and Philippines (FAO, 2016). Expansion of plantation crops such as rubber (Fox and Castella, 2013), oil palm (Castellanos-Navarrete and Jansen, 2015), and sugarcane threaten small-holder farms and reduce land available to local food production.

5.1. Classification

Previous studies have mapped cropland extent in Southeast Asia, but they have been limited by low spatial resolution or low spatial extent. This study covered all of Southeast Asia over the years between 2013–2016, using Landsat 30-m imagery. Landsat 8 and 7 were used to maximize data coverage to every 8 days over 3 years (2013–2016). Time-compositing (Fig. 3) over such a dense time-series, in all four time-periods (Fig. 2), ensured that less than 1% of pixels were obscured by clouds across the entire Southeast and Northeast Asia study area (Fig. 1) (in Borneo the amount may be closer to 2–3%). Further, these pixels are in deep forests, having inconsequential impact on cropland class accuracies. With the increasing availability of Sentinel 2 imagery, its use with Landsat should be considered, because the increase in temporal acquisition outweighs drawbacks from mismatch using multiple sensors (Xiong et al., 2017b). In particular, the development of Sentinel 2 – Landsat harmonized datasets makes it easier to merge Landsat and Sentinel 2 imagery (Claverie et al., 2017). Increasing temporal acquisition from multiple satellites also increases the performance of data interpolation and curve fitting algorithms (Brooks et al., 2012; Nguyen et al., 2012; Vuolo et al., 2017). Additionally, Landsat 9, which is similar to Landsat 8, is currently being constructed and is on track to be launched in 2020, ensuring the continuity and relevance of Landsat imagery (Jenstrom and Sauer, 2018).

Google Earth Engine was used to generate cloud-free image

Table 5

Overall accuracy error matrix of the Southeast and Northeast Asian countries (SNAC) study area which includes 12 nations. Independently collected validation samples were used to assess the accuracy of the cropland product and show the overall, producer's and user's accuracies. TNCA = Total Net Cropland Area of SNAC.

Entire Study Area: 100 % of TNCA				
	Cropland	Non cropland	Row total	Commission error
Cropland	376	114	490	23.3%
Non cropland	85	1175	1260	6.7%
Column total	461	1289	1750	
Omission error	18.4%	8.8%		
Producer accuracy	81.6%	91.2%		
User accuracy	76.7%	93.3%		
Overall accuracy				88.6%

Table 6

Accuracy error matrices of the 7 refined agro-ecological zones (RAEZs) of Southeast and Northeast Asian countries (SNAC) study area. Independent assessment of overall, producer's, and user's accuracies are shown for the circa 2015 cropland extent product derived from Landsat imagery. TNCA = Total Net Cropland Area of SNAC.

RAEZ 1: Mainland SE Asia: 47.6 % of TNCA				
	Cropland	Non cropland	Row total	Commission error
Cropland	80	20	100	20.0%
Non cropland	16	134	150	10.7%
Column total	96	154	250	
Omission error	16.7%	13.0%		
Producer accuracy	83.3%	87.0%		
User accuracy	80.0%	89.3%		
Overall accuracy				85.6%
RAEZ 2: Philippines: 7.2 % of TNCA				
	Cropland	Non cropland	Row total	Commission error
Cropland	41	7	48	14.6%
Non cropland	15	187	202	7.4%
Column total	56	194	250	
Omission error	26.8%	3.6%		
Producer accuracy	73.2%	96.4%		
User accuracy	85.4%	92.6%		
Overall accuracy				91.2%
RAEZ 3: Sumatra & W. Malaysia: 14.5 % of TNCA				
	Cropland	Non cropland	Row total	Commission error
Cropland	45	20	65	30.8%
Non cropland	22	163	185	11.9%
Column total	67	183	250	
Omission error	32.8%	10.9%		
Producer accuracy	67.2%	89.1%		
User accuracy	69.2%	88.1%		
Overall accuracy				83.2%
RAEZ 4: Java & Bali: 4.7 % of TNCA				
	Cropland	Non cropland	Row total	Commission error
Cropland	89	27	116	23.3%
Non cropland	13	121	134	9.7%
Column total	102	148	250	
Omission error	12.7%	18.2%		
Producer accuracy	87.3%	81.8%		
User accuracy	76.7%	90.3%		
Overall accuracy				84.0%
RAEZ 5: Borneo: 14.4 % of TNCA				
	Cropland	Non cropland	Row total	Commission error
Cropland	69	26	95	27.4%
Non cropland	2	153	155	1.3%
Column total	71	179	250	
Omission error	2.8%	14.5%		
Producer accuracy	97.2%	85.5%		
User accuracy	72.6%	98.7%		
Overall accuracy				88.8%
RAEZ 6: Japan and N&S Korea: 6.7 % of TNCA				
	Cropland	Non cropland	Row total	Commission error
Cropland	42	7	49	14.3%
Non cropland	15	186	201	7.5%
Column total	57	193	250	
Omission error	26.3%	3.6%		
Producer accuracy	73.7%	96.4%		
User accuracy	85.7%	92.5%		
Overall accuracy				91.2%

Table 6 (continued)

RAEZ 7: Eastern Indonesian Islands: 4.9 % of TNCA				
	Cropland	Non cropland	Row total	Commission error
Cropland	10	7	17	41.2%
Non cropland	2	231	233	0.9%
Column total	12	238	250	
Omission error	16.7%	2.9%		
Producer accuracy	83.3%	97.1%		
User accuracy	58.8%	99.1%		
Overall accuracy				96.4%

composites from Landsat 8 and 7 imagery. Through experimentation, we found that it was necessary to make composites covering 4 months and over 3 years (Fig. 2 and 3) in order to create consistent cloud-free composites over all of the study area. Additionally, we observed that including Landsat 7 (despite issues with scan lines) with Landsat 8 produced better composites than Landsat 8 alone. In order to not remove any potentially useful information, all Landsat 7 bands comparable to the Landsat 8 bands (except the 15 m Pan) were used for classification (Table 2). In addition, three spectral indices NDVI, NBR2, and LSWI (Section 3.2) were included to help classify vegetation, bare ground, and surface water respectively. Additionally, slope and elevation, derived from SRTM 30 m, were included since precipitation, temperature, land cover, and agricultural practices vary with slope and elevation.

The band values for each sample in a RAEZ were plotted in a bar and whisker plot to evaluate the distribution and separability (Fig. 5). None of the 30 4-month-period composite bands demonstrated a clear separability between cropland and non-croplands. We hypothesized that the broad composites masked the growth stages (planting, growing, fruiting, harvest) of seasonal crops which distinguish themselves from other types of land cover. To reintroduce the spectral variability from those different growth stages, the standard deviation for all cloud free images within the year 2015 were also included in the classification. Fig. 5 demonstrated that the standard deviation band had greater class separability for a given band than any one seasonal composite for that band.

The RF algorithm was chosen for the map classification because it has been used successfully for over a decade in remote landscape classifications, and is resistant to highly correlated data and for over-coming data over-fitting (Belgiu and Drăgu, 2016). As a result, high band correlation was not a concern for decreasing classifier performance. Analysts selected 90-m x 90-m areas for training data. A minimum of 500 training samples per RAEZ were generated and used. An iterative process of classification, review, adding samples in misclassified areas, and reclassification was performed to improve the product.

5.2. Cropland class accuracies and uncertainties

Often when performing land cover classifications, there is a trade-off in increasing accuracies: when producer's accuracies increase, user's accuracies decrease and vice versa. The RF algorithm is optimized to keep both accuracies as high as possible. However, when optimizing the RF algorithm, our goal was to capture as much cropland as possible (i.e., to keep the errors of omissions minimum or producer's accuracies high). In the process, if some non-croplands are classified as croplands, the errors of commissions go up (or user's accuracies go down). The mixed pixel challenge using 30 m (1 pixel = 0.09 ha) Landsat data is minimum compared to much coarser resolution data such as MODIS 250 m (1 pixel = 6.26 ha). This is because the purity of the training and validation data for 30-m pixels are significantly higher than much

Table 7

Comparison of cropland areas derived from this product (i.e., Global Food Security-support Analysis Data @ 30-m for Southeast Asia (GFSAD30SEA)) with: MIRCA2000 (personal communication with Stefan Siebert, 2014); FAO cultivated area (2009); GRIPC (Salmon et al., 2015); GIAM (Thenkabail et al., 2009) and GMRCA (Biradar et al., 2009). All cropland product definitions are the same except GRIPC where cropland fallows are not included^{***}. All areas are listed in thousands of ha (x 1000 ha).

Country Resolution	Total Land Area GAO-GAUL	GFSAD30SEACE This Study: 30 m	Crop Extent MIRCA variable	FAO 2009 Cultivated Area variable	GRIPC 500 m	GIAM + GMRCA 1 km
Indonesia	181,100	37,440	55,750	42,600	31,030	20,750
Thailand	51,150	25,760	18,320	19,000	27,360	16,540
Myanmar	65,500	14,240	12,640	12,130	11,560	10,710
Vietnam	31,000	10,800	9,520	9,640	12,060	10,350
Malaysia	32,800	10,420	7,770	7,590	4,700	5,300
Philippines	29,800	9,150	12,190	10,440	12,820	9,020
Cambodia	17,600	7,680	3,980	4,060	4,980	4,600
Japan	36,580	3,730	5,790	4,610	5,990	5,950
North Korea	12,050	3,330	3,060	2,860	5,250	3,070
Laos	23,000	2,450	1,110	1,470	1750	2,020
South Korea	9,7000	1,520	2,090	1,780	3,180	3,120
Brunei	577	54	13	11	0	
Total	490,857	126,574	132,233	116,191	120,680	91,430

Sources.

MIRCA: MIRCA2000-Global Monthly Irrigated and Rainfed Crop Areas around the year 2000.

FAO Cultivated Area circa 2009: FAO Yearbook 2013.

GRIPC: Global Rain-fed, Irrigated, & Paddy Croplands.

GIAM: Global Irrigated Area Map derived from remote sensing, for the end of the last millennium.

GMRCA: Global Map of Rainfed Cropland Areas around the year 2000.

*** = definition of croplands for different products.

GFSAD30SEACE = standing crops + cropland fallows + permanent crops.

MIRCA = Irrigated + rainfed annual and perennial cropland + temporary fallows.

FAO 2009 = Annual crops + perennial crops + cropland fallows.

GRIPC = Irrigated, rainfed, and paddy (rice) cropland that are harvested at least once per year (fallow not included).

GIAM + GMRA = standing crops + cropland fallows + permanent crops.

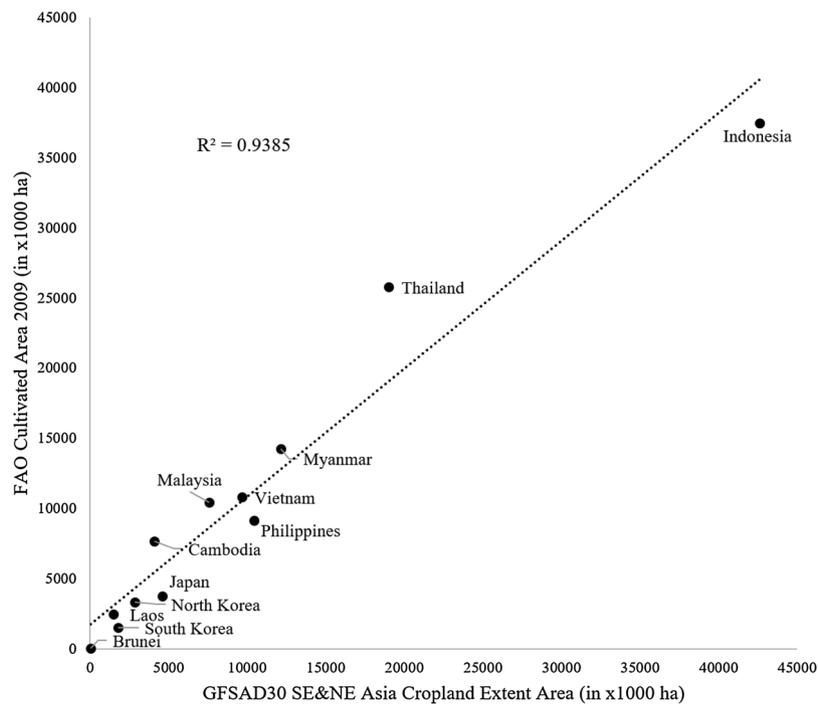


Fig. 8. Comparison of Landsat 30-m cropland areas derived from this study with UN FAO cultivated area national statistics for the 12 Southeast and Northeast Asian countries.

coarser pixels such as MODIS 250 m. This leads to substantially reduced uncertainties in the random forest classifier as well as reduction in uncertainties in accuracy assessments of a 30-m data as opposed to much coarser data such as MODIS 250 m.

5.2.1. Cropland over-classification

In RAEZ 1 (Mainland Southeast Asia), cropland was slightly over-classified. This finding is evident with both the producer's accuracy being higher than the user's accuracy and the cropland extent calculated from GFSAD30 being higher for Thailand, Vietnam, Cambodia, Laos, and Myanmar than that reported in FAO statistics and many of the other cropland extent datasets (Table 7).

The accuracy assessment of RAEZ 5 (Borneo) suggests that cropland was over-classified. The producer's accuracy was significantly higher than the user's accuracy for the cropland class, 97.2% versus 72.6% respectively, suggesting somewhat higher cropland commissions. Borneo has particularly heavy cloud cover due to heavy and regular rainfall (see Table 1). Additionally, peatland fires regularly occur, producing large amounts of smoke (Miettinen et al., 2016; Uda et al., 2017). The combination of heavy cloud cover and smoke reduced the quality of the image composites in this RAEZ more than any other, which likely negatively impacted classification accuracy.

5.2.2. Cropland under-classification

When classifying RAEZ 2 (Philippines) it was particularly difficult to classify plantations as cropland without classifying large sections of forest as cropland as well. Additionally, in many areas of the Philippines, there is a checkerboard like pattern of croplands and forest, with agricultural fields less than 1 ha in size, which are difficult to classify. This corresponds with the under-estimation of agricultural area in the Philippines compared to other studies as GFSAD30 found 9,150,000 ha, whereas the FAO (which reports data from the Agricultural Ministry of Philippines) listed 10,440,000 ha. Cropland area was also under-classified in RAEZ 6 (Japan and N&S Korea). Upon reviewing the product on www.croplands.org, it was found that very small farms were sometimes omitted. Particularly in Japan, it is common for very small fields (> 1 ha) to be located among houses on the edges of cities. Some of these fields were not classified as cropland, largely due to mixed pixels (cropland and developed land occurring within the same pixel). This finding corresponds with some under-estimation of agricultural area in Japan compared to other studies as GFSAD30 found 3,730,000 ha, whereas FAO (which used data from the Agricultural Ministry of Japan) reported 4,610,000 ha.

5.3. Sub-national cropland area comparisons

A primary benefit of mapping cropland extent at 30 m is for assessing agricultural land at the regional or local level. Mapping croplands at 30 m delineates field boundaries and can capture small isolated farms. To assess how well GFSAD30SEACE (this study) mapped cropland at the provincial level, our product was compared to cropland extent statistics obtained from respective countries' Ministry of Agriculture or Statistics where such data were available.

For Thailand, the over-estimation of cropland extent at the Provincial scale (Fig. 10) was based on the linear relationship between GFSAD30 and the Whole Kingdom of Thailand 2013 Agricultural Census (National Statistical Office, 2013). This over-estimation is understandable considering the commissions (classifying non-cropland area as cropland) is preferable to omissions (classifying cropland as non-cropland). Regardless, the high R^2 value of 0.948 of the linear fit indicates consistent cropland mapping across the entire country of Thailand.

For South Korea, the Agricultural Land Area Survey provided by the Korean Statistical Information Service (Statistics Korea, 2017) was in close agreement with the cropland areas obtained from GFSAD30-SEACE derived 30-m statistics of this study (Fig. 10). The high R^2 value

of 0.952 of the linear fit between GFSAD30 and the provincial areas indicates consistent cropland mapping across the entire country.

6. Conclusions

This study successfully produced high spatial resolution 30-m (1 pixel = 0.09 ha) cropland extent product for the entire 12 Southeast and Northeast Asian Countries (SNAC). Overall, 25.78% (126,574,000 ha) of the total geographic area was net cropland area for the nominal year 2015. In comparison, only 12.6% of the terrestrial land in the world is under croplands. This shows the intense cropland activity in Southeast and Northeast Asia as a result of favourable climate, adequate precipitation and/or irrigation, rich soils, and the need to feed a population 836.7 million from the 12 study countries. The product has high robust levels of accuracy, established by an independent group, with a weighted overall accuracy of 88.6%, a producer's accuracy of 81.6% (error of omission = 18.4%) and a user's accuracy of 76.7% (error of commission = 23.3%). For each of the 7 RAEZs overall accuracies varied from 83.2 to 96.4%. The computed cropland areas of the countries and sub-national levels explained 93–95% of the variability when compared with areas reported by the UN FAO and the national statistics. The uniqueness of this study was clearly demonstrated by the applicability of the pixel-based random forest machine learning algorithm over such large areas using knowledge generated by well distributed, large sample size reference training data through petabyte-scale powerful cloud computing on Google Earth Engine. The cropland product can be viewed at full resolution at www.croplands.org and downloaded from the NASA LP DAAC at: <https://lpdaac.usgs.gov/node/1281>.

Acknowledgements

This project was funded by NASA MEaSUREs, through the NASA ROSES solicitation, and funded for a period of 5 years (June 1, 2013–May 31, 2018). The NASA Making Earth System Data Records for Use in Research Environments (MEaSUREs) grant number is NNH13AV82I and the USGS Sales Order number is 29039. We gratefully acknowledge this support. The U.S. Geological Survey (USGS) provided significant direct and indirect supplemental funding through the Land Resources Mission Area (LRMA), National Land Imaging (NLI) Program, and Land Change Science (LCS) program. We gratefully acknowledge this support. This research is a part of the Global Food Security -support Analysis Data Project at 30-m (GFSAD30). The project was led by the USGS in collaboration with NASA Ames, University of New Hampshire (UNH), University of Wisconsin (UW), NASA Goddard Space Flight Center (GSFC), and the Northern Arizona University (NAU). We also collaborated with international partners including The International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) and The Asian Disaster Preparedness Center (ADPC). The authors would like to thank Dr. Felix T. Portman and Dr. Stefan Siebert for providing statistics of MIRCA2000 (Portmann et al., 2010; also, latest statistics through personal communication between. Dr. Stefan Siebert and Prasad S. Thenkabail) for inter-comparison. Finally, special thanks to Ms. Susan Benjamin, Director of the USGS Western Geographic Science Center (WGSC) and Mr. Larry Gaffney, Administrative Officer of WGSC for their support and encouragement throughout this research. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

References

- Agri-Food & Veterinary Authority of Singapore, 2018. Farming in Singapore. [WWW Document]. URL: <http://www.ava.gov.sg/explore-by-sections/farms/land-farms/farming-in-singapore>.
- Ahamed, A., Bolten, J.D., 2017. A MODIS-based automated flood monitoring system for southeast asia. *Int. J. Appl. Earth Obs. Geoinf.* 61, 104–117. <https://doi.org/10.1016/j.jag.2017.05.006>.

- Arino, O., Gross, D., Ranera, F., Leroy, M., Bicheron, P., Brockman, C., Defourny, P., Vancutsem, C., Achard, F., Durieux, L., Bourg, L., Latham, J., Di Gregorio, A., Witt, R., Herold, M., Sambale, J., Plummer, S., Weber, J.-L., 2007. GlobCover: ESA service for global land cover from MERIS. 2007 IEEE International Geoscience and Remote Sensing Symposium. IEEE 2412–2415. <https://doi.org/10.1109/IGARSS.2007.4423328>.
- Asian Development Bank, 2009. The economics of climate change in Southeast Asia: a regional review. *Asian Dev. Bank* 255.
- Bartholomé, E., Belward, A.S., 2005. GLC2000: a new approach to global land cover mapping from Earth observation data. *Int. J. Remote Sens.* 26, 1959–1977. <https://doi.org/10.1080/01431160412331291297>.
- Beckschäfer, P., 2017. Obtaining rubber plantation age information from very dense Landsat TM & ETM + time series data and pixel-based image compositing. *Remote Sens. Environ.* 196, 89–100. <https://doi.org/10.1016/j.rse.2017.04.003>.
- Belgiu, M., Drăgu, L., 2016. Random forest in remote sensing: a review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* 114, 24–31. <https://doi.org/10.1016/j.isprsjprs.2016.01.011>.
- Bicheron, P., Amberg, V., Bourg, L., Petit, D., Huc, M., Miras, B., Brockmann, C., Hagolle, O., Delwart, S., Ranera, F., Leroy, M., Arino, O., 2011. Geolocation assessment of MERIS GlobCover orthorectified products. *IEEE Trans. Geosci. Remote Sens.* 49, 2972–2982. <https://doi.org/10.1109/TGRS.2011.2122337>.
- Biradar, C.M., Thenkabail, P.S., Noolipady, P., Li, Y., Dheeravath, V., Turral, H., Velpuri, M., Gumma, M.K., Gangalakunta, O.R.P., Cai, X.L., Xiao, X., Schull, M.A., Alankara, R.D., Gunasinghe, S., Mohideen, S., 2009. A global map of rainfed cropland areas (GMRFCA) at the end of last millennium using remote sensing. *Int. J. Appl. Earth Obs. Geoinf.* 11, 114–129. <https://doi.org/10.1016/j.jag.2008.11.002>.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45, 5–32. <https://doi.org/10.1023/A:1010933404324>.
- Breiman, L., Cutler, A., 2004. Random Forests. [WWW Document]. Berkeley. URL: https://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm.
- Bren d'Amour, C., Reitsma, F., Baiocchi, G., Barthel, S., Güneralp, B., Erb, K.-H., Haberl, H., Creutzig, F., Seto, K.C., 2016. Future urban land expansion and implications for global croplands. *Proc. Natl. Acad. Sci.* 201606036. <https://doi.org/10.1073/pnas.1606036114>.
- Brooks, E.B., Thomas, V.A., Wynne, R.H., Coulston, J.W., 2012. Fitting the multipletoral curve: a fourier series approach to the missing data problem in remote sensing analysis. *IEEE Trans. Geosci. Remote Sens.* 50, 3340–3353. <https://doi.org/10.1109/TGRS.2012.2183137>.
- Brown, M.E., de Beurs, K.M., Marshall, M., 2012. Global phenological response to climate change in crop areas using satellite remote sensing of vegetation, humidity and temperature over 26years. *Remote Sens. Environ.* 126, 174–183. <https://doi.org/10.1016/j.rse.2012.08.009>.
- Bruinsma, J., 2009. By how much do land, water and crop yields need to increase by 2050? The resource outlook to 2050. Food and Agriculture Organization of the United Nations Expert Meeting Entitled “How to Feed the World In. pp. 24–26.
- Bui, T.D., Maier, S.W., Austin, C.M., 2013. Land cover and land use change related to shrimp farming in coastal areas of Quang Ninh, Vietnam using remotely sensed data. *Environ. Earth Sci.* 72, 441–455. <https://doi.org/10.1007/s12665-013-2964-0>.
- Castellanos-Navarrete, A., Jansen, K., 2015. Oil palm expansion without enclosure: smallholders and environmental narratives. *J. Peasant Stud.* 42, 791–816. <https://doi.org/10.1080/03066150.2015.1016920>.
- Chen, J., Chen, J., Liao, A., Cao, X., Chen, L., Chen, X.H., He, C., Han, G., Peng, S., Lu, M., Zhang, W., Tong, X., Mills, J., 2015. Global land cover mapping at 30m resolution: a POK-based operational approach. *ISPRS J. Photogramm. Remote Sens.* 103, 7–27. <https://doi.org/10.1016/j.isprsjprs.2014.09.002>.
- Chen, Y., Lu, D., Moran, E., Batistella, M., Dutra, L.V., Sanches, I.D., da Silva, R.F.B., Huang, J., Luiz, A.J.B., de Oliveira, M.A.F., 2018. Mapping croplands, cropping patterns, and crop types using MODIS time-series data. *Int. J. Appl. Earth Obs. Geoinf.* 69, 133–147. <https://doi.org/10.1016/j.jag.2018.03.005>.
- Cheng, Y., Yu, L., Cracknell, A.P., Gong, P., 2016. Oil palm mapping using Landsat and PALSAR: a case study in Malaysia. *Int. J. Remote Sens.* 37, 5431–5442. <https://doi.org/10.1080/01431161.2016.1241448>.
- Claverie, M., Masek, J.G., Ju, J., Dungan, J.L., 2017. Harmonized Landsat-8 Sentinel-2 (HLS) Product User's Guide Version: 1.3.
- Congalton, R.G., Kass, G., 2008. Assessing the Accuracy of Remotely Sensed Data: Principles and Practices. CRC Press/Taylor & Francis. <https://doi.org/10.1111/j.1477-9730.2010.00574.2.x>.
- Congalton, G.R., Yadav, K., McDonnell, K., Poehnel, J., Stevens, B., Gumma, K.M., Teluguntla, P., 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 2015 Validation Global 30 m V001 [Data set]. NASA EOSDIS Land Processes DAAC. <https://doi.org/10.5067/MEaSUREs/GFSAD/GFSAD30VAL.001>.
- Dong, J., Xiao, X., 2016. Evolution of regional to global paddy rice mapping methods: a review. *ISPRS J. Photogramm. Remote Sens.* 119, 214–227. <https://doi.org/10.1016/j.isprsjprs.2016.05.010>.
- Dutta, R., Basnayake, S., Ahmed, A.K., 2015. Assessing gaps and strengthening early warning system to manage disasters in Cambodia. *J. Integr. Disaster Risk Manag.* 5. <https://doi.org/10.5595/idrim.2015.0104>.
- FAO, 2013. Fao Statistical Yearbook 2013. [WWW Document]. Crop Prod. Stat. Food Agric. Organ, Rome. <https://doi.org/10.1017/CBO9781107415324.004>.
- FAO, 2016. Regional Overview of Food Insecurity: Asia and the Pacific. Bangkok. <https://doi.org/ISBN978-92-5-109514-0>.
- Farr, T.G., Rosen, P.A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D., Alsdorf, D.E., 2007. The shuttle radar topography mission. *Rev. Geophys.* 45, RG2004. <https://doi.org/10.1029/2005RG000183>.
- Fischer, G., van Velthuizen, H.T., Nachtergaele, F.O., 2000. Global Agro-ecological Zones Assessment: Methodology and Results. Food and Agriculture Organization of the United Nations, Rome.
- Fox, J., Castilla, J.C., 2013. Expansion of rubber (Hevea brasiliensis) in Mainland Southeast Asia: What are the prospects for smallholders? *J. Peasant Stud.* 40, 155–170. <https://doi.org/10.1080/03066150.2012.750605>.
- Friedl, M.A., McIver, D.K., Hodges, J.C.F., Zhang, X.Y., Muchoney, D., Strahler, A.H., Woodcock, C.E., Gopal, S., Schneider, A., Cooper, A., Baccini, A., Gao, F., Schaaf, C., 2002. Global land cover mapping from MODIS: algorithms and early results. *Remote Sens. Environ.* 83, 287–302. [https://doi.org/10.1016/S0034-4257\(02\)00078-0](https://doi.org/10.1016/S0034-4257(02)00078-0).
- Gallego, F.J., Kussul, N., Skakun, S., Kravchenko, O., Shelestov, A., Kussul, O., 2014. Efficiency assessment of using satellite data for crop area estimation in Ukraine. *Int. J. Appl. Earth Obs. Geoinf.* 29, 22–30. <https://doi.org/10.1016/j.jag.2013.12.013>.
- Gerbens-Leenes, P.W., Nonhebel, S., 2004. Critical water requirements for food, methodology and policy consequences for food security. *Food Policy* 29, 547–564. <https://doi.org/10.1016/j.foodpol.2004.09.003>.
- Giri, C., Defourny, P., Shrestha, S., 2003. Land cover characterization and mapping of continental Southeast Asia using multi-resolution satellite sensor data. *Int. J. Remote Sens.* 24, 4181–4196. <https://doi.org/10.1080/0143116031000139827>.
- 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.* <https://doi.org/10.1016/j.rse.2017.06.031>.
- Gumma, M.K., Thenkabail, P.S., Teluguntla, P., Rao, M.N., Mohammed, I.A., 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. *Int. J. Digit. Earth* 9, 981–1003. <https://doi.org/10.1080/17538947.2016.1168489>.
- Gumma, M.K., Thenkabail, P.S., Deevi, K.C., Mohammed, I.A., Teluguntla, P., Oliphant, A.J., Xiong, J., Aye, T., Whitbread, A.M., 2018. Mapping cropland fallow areas in myanmar to scale up sustainable intensification of pulse crops in the farming system. *GIScience Remote Sens.* 12, 1–24. <https://doi.org/10.1080/15481603.2018.1482855>.
- Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S., Tyukavina, A., Thau, D., Stehman, S.V., Goetz, S.J., Loveland, T.R., Kommareddy, A., Egorov, A., Chini, L., Justice, C.O., Townshend, J.R., 2013. High-resolution global maps of 21st-Century forest cover change. *Science* 342 (80-), 850–853. <https://doi.org/10.1126/science.1244693>.
- 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 (No. RSAC-10108-RPT1). Salt Lake City.
- Hurni, K., Schneider, A., Heinimann, A., Nong, D.H., Fox, J., 2017. Mapping the expansion of boom crops in Mainland Southeast Asia using dense time stacks of landsat data. *Remote Sens.* 9, 320. <https://doi.org/10.3390/rs9040320>.
- Iizumi, T., Ramankutty, N., 2015. How do weather and climate influence cropping area and intensity? *Glob. Food Sec.* 4, 46–50. <https://doi.org/10.1016/j.gfs.2014.11.003>.
- Jenstrom, D., Sauer, B., 2018. Landsat 9 Project Status, in: Landsat Science Team Meeting. US Geological Survey, Sioux Falls, SD.
- Kobrick, M., Crippen, R., 2014. SRTMIMG: NASA Shuttle Radar Topography Mission Combined Image Data Set V003. NASA EOSDIS Land Processes DAAC. <https://doi.org/10.5067/MEaSUREs/SRTM/SRTMIMG.003>.
- Kontgis, C., Schneider, A., Ozdogan, M., 2015. Mapping rice paddy extent and intensification in the Vietnamese Mekong River Delta with dense time stacks of Landsat data. *Remote Sens. Environ.* 169, 255–269. <https://doi.org/10.1016/j.rse.2015.08.004>.
- Latham, J., Cumani, R., Rosati, I., Bloise, M., 2014. Global Land Cover SHARE (GLC-SHARE) Database Beta-Release Version 1.0-2014. Food Agric. Organ. UN,....
- Lee, J.S.H., Wich, S.A., Widayati, A., Koh, L.P., 2016. Detecting industrial oil palm plantations on Landsat images with Google Earth Engine. *Remote Sens. Appl. Soc. Environ.* 4, 219–224. <https://doi.org/10.1016/j.rsae.2016.11.003>.
- Li, B.V., Hughes, A.C., Jenkins, C.N., Ocampo-Peñuela, N., Pimm, S.L., 2016. Remotely sensed data informs Red List evaluations and conservation priorities in Southeast Asia. *PLoS One* 11, e0160566. <https://doi.org/10.1371/journal.pone.0160566>.
- Li, P., Feng, Z., Xiao, C., 2017. Acquisition probability differences in cloud coverage of the available Landsat observations over mainland Southeast Asia from 1986 to 2015. *Int. J. Digit. Earth* 1–14. <https://doi.org/10.1080/17538947.2017.1327619>.
- Manjunath, K.R., More, R.S., Jain, N.K., Panigrahy, S., Parihar, J.S., 2015. Mapping of rice-cropping pattern and cultural type using remote-sensing and ancillary data: a case study for South and Southeast Asian countries. *Int. J. Remote Sens.* 36, 6008–6030. <https://doi.org/10.1080/01431161.2015.1110259>.
- Meybeck, A., Lankoski, J., Redfern, S., Azzu, N., Gitz, V., 2012. FAO, RomeBuilding Resilience for Adaptation to Climate Change in the Agriculture Sector, Proceedings of a Joint FAO/OECD Workshop 23–24 April 2012. <https://doi.org/ISBN978-92-5-107373-5>.
- Miettinen, J., Shi, C., Liew, S.C., 2016. Land cover distribution in the peatlands of Peninsular Malaysia, Sumatra and Borneo in 2015 with changes since 1990. *Glob. Ecol. Conserv.* 6, 67–78. <https://doi.org/10.1016/j.gecco.2016.02.004>.
- Monfreda, C., Ramankutty, N., Foley, J.A., 2008. Farming the planet: 2. Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000. *Global Biogeochem. Cycles* 22, GB1022. <https://doi.org/10.1029/2007GB002947>.
- Mutert, B.E., Fairhurst, T.H., 2002. Developments in rice production in Southeast Asia. *Better Crop. Int.* 15, 12–17.
- National Statistical Office, 2013. Whole Kingdom of Thailand 2013 Agricultural Census. Bangkok.
- Nguyen, T.T.H., De Bie, C.A., Ali, A., Smaling, E.M.A., Chu, T.H., 2012. Mapping the

- irrigated rice cropping patterns of the Mekong delta, Vietnam, through hyper-temporal SPOT NDVI image analysis. *Int. J. Remote Sens.* 33, 415–434. <https://doi.org/10.1080/01431161.2010.532826>.
- Oliphant, A.J., Li, J., Wynne, R.H., Donovan, P.F., Zipper, C.E., 2014. Identifying woody vegetation on coal surface mines using phenological indicators with multitemporal landsat imagery. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives* 339–345. <https://doi.org/10.5194/isprsarchives-XL-1-339-2014>.
- Oliphant, A.J., Thenkabail, P.S., Teluguntla, P., Xiong, J., Congalton, R.G., Yadav, K., Massey, R., Gumma, M.K., Smith, C., 2017a. NASA Making Earth System Data Records for Use in Research Environments (MEAsURES) Global Food Security-support Analysis Data (GFSAD) @ 30-m for Southeast & Northeast Asia: Cropland Extent Product (GFSAD30SEACE). [WWW Document]. NASA EOSDIS L. Process. DAAC. <https://doi.org/10.5067/MEAsURES/GFSAD/GFSAD30SEACE.001>.
- Oliphant, A.J., Wynne, R.H., Zipper, C.E., Ford, W.M., Donovan, P.F., Li, J., 2017b. Autumn olive (*Elaeagnus umbellata*) presence and proliferation on former surface coal mines in Eastern USA. *Biol. Invasions* 19, 179–195. <https://doi.org/10.1007/s10530-016-1271-6>.
- Panuju, D.R., Mizuno, K., Trisasongko, B.H., 2013. The dynamics of rice production in Indonesia 1961–2009. *J. Saudi Soc. Agric. Sci.* 12, 27–37. <https://doi.org/10.1016/j.jssas.2012.05.002>.
- Peel, M.C., Finlayson, B.L., McMahon, T.A., 2007. Updated world map of the Köppen-Geiger climate classification. *Hydrol. Earth Syst. Sci.* 11, 1633–1644. <https://doi.org/10.5194/hess-11-1633-2007>.
- Pelletier, C., Valero, S., Inglada, J., Champion, N., Dedieu, G., 2016. Assessing the robustness of Random Forests to map land cover with high resolution satellite image time series over large areas. *Remote Sens. Environ.* 187, 156–168. <https://doi.org/10.1016/j.rse.2016.10.010>.
- 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. <https://doi.org/10.1029/2008GB003435>. n/a-n/a.
- Reda, A.G., Tripathi, N.K., 2016. *Climate Dynamics and Rice Agriculture in SE Asia*. Abbadji Girmay Reda and Nitin K. Tripathi. OMICS Group eBooks.
- Reid, J.S., Hyer, E.J., Johnson, R.S., Holben, B.N., Yokelson, R.J., Zhang, J., Campbell, J.R., Christopher, S.A., Di Girolamo, L., Giglio, L., Holz, R.E., Kearney, C., Miettinen, J., Reid, E.A., Turk, F.J., Wang, J., Xian, P., Zhao, G., Balasubramanian, R., Chew, B.N., Janjai, S., Lagrosas, N., Lestari, P., Lin, N.H., Mahmud, M., Nguyen, A.X., Norris, B., Oanh, N.T.K., Oo, M., Salinas, S.V., Welton, E.J., Liew, S.C., 2013. Observing and understanding the Southeast Asian aerosol system by remote sensing: an initial review and analysis for the seven Southeast Asian Studies (7SEAS) program. *Atmos. Res.* <https://doi.org/10.1016/j.atmosres.2012.06.005>.
- Rodriguez-Galiano, V.F., Ghimire, B., Rogan, J., Chica-Olmo, M., Rigol-Sanchez, J.P., 2012. An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS J. Photogramm. Remote Sens.* 67, 93–104. <https://doi.org/10.1016/j.isprsjprs.2011.11.002>.
- Salmon, J.M., Friedl, M.A., Frolking, S., Wisser, D., Douglas, E.M., 2015. Global rain-fed, irrigated, and paddy croplands: a new high resolution map derived from remote sensing, crop inventories and climate data. *Int. J. Appl. Earth Obs. Geoinf.* 38, 321–334. <https://doi.org/10.1016/j.jag.2015.01.014>.
- See, L., Fritz, S., You, L., Ramankutty, N., Herrero, M., Justice, C.O., Becker-Reshef, I., Thornton, P., Erb, K.-H., Gong, P., Tang, H., van der Velde, M., Ericksen, P., McCallum, I., Kraxner, F., Obersteiner, M., 2015. Improved global cropland data as an essential ingredient for food security. *Glob. Food Sec.* 4, 37–45. <https://doi.org/10.1016/j.gfs.2014.10.004>.
- Seong, J.C., Choi, J., Choi, Y.S., Yoon, H., 2008. Applications of point-to-point distances with regional and global datasets. *GIScience Remote Sens.* 45, 369–376. <https://doi.org/10.2747/1548-1603.45.3.369>.
- Setiawan, Y., Yoshino, K., Philpot, W.D., 2013. Characterizing temporal vegetation dynamics of land use in regional scale of Java Island, Indonesia. *J. Land Use Sci.* 8, 1–30. <https://doi.org/10.1080/1747423X.2011.605178>.
- Sharma, R.C., Tateishi, R., Hara, K., Iizuka, K., 2016. Production of the Japan 30-m land cover map of 2013–2015 using a random forests-based feature optimization approach. *Remote Sens.* 8, 429. <https://doi.org/10.3390/rs8050429>.
- Sianturi, R.S., Jetten, V.G., Sartohadi, J., 2018. Mapping cropping patterns in irrigated rice fields in West Java: towards mapping vulnerability to flooding using time-series MODIS imagery. *Int. J. Appl. Earth Obs. Geoinf.* 66, 1–13. <https://doi.org/10.1016/j.jag.2017.10.013>.
- Sonobe, R., Yamaya, Y., Tani, H., Wang, X., Kobayashi, N., Mochizuki, Kichiro, 2017. Assessing the suitability of data from Sentinel-1A and 2A for crop classification. *GIScience Remote Sens.* 54, 918–938. <https://doi.org/10.1080/15481603.2017.1351149>.
- Statistics Korea, 2017. *Agricultural Land Area*. [WWW Document]. URL. http://kosis.kr/statHtml/statHtml.do?orgId=101&tblId=DT_1EB001&language=en&conn_path=13.
- Suepa, T., Qi, J., Lawawirojwong, S., Messina, J.P., 2016. Understanding spatio-temporal variation of vegetation phenology and rainfall seasonality in the monsoon Southeast Asia. *Environ. Res.* 147, 621–629. <https://doi.org/10.1016/j.envres.2016.02.005>.
- Teluguntla, P., Thenkabail, P.S., Xiong, J., Gumma, M.K., Giri, C., Milesi, C., Ozdogan, M., Congalton, R.G., Tilton, J.C., Sankey, T.T., Massey, R., Yadav, K., 2016. 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.), *Land Resources Monitoring, Modeling, and Mapping With Remote Sensing*. CRC Press, pp. 131–160.
- Teluguntla, P., Thenkabail, P.S., Xiong, J., Gumma, M.K., Congalton, R.G., Oliphant, A.J., Poehnel, J., Yadav, K., Rao, M., 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. *Int. J. Digit. Earth* 10, 944–977. <https://doi.org/10.1080/17538947.2016.1267269>.
- Teluguntla, P., Thenkabail, P.S., Oliphant, A.J., Xiong, J., Gumma, M.K., Congalton, R.G., Yadav, K., Huete, A., 2018. A 30-m landsat-derived cropland extent product of Australia and China using random forest machine learning algorithm on google earth engine cloud computing platform. *ISPRS J. Photogramm. Remote Sens.* 144, 325–340. <https://doi.org/10.1016/j.isprsjprs.2018.07.017>.
- Thenkabail, P.S., Biradar, C.M., Noojipady, P., Cai, X., Dheeravath, V., Li, Y., Velpuri, M., Gumma, M.K., Pandey, S., 2007. Sub-pixel area calculation methods for estimating irrigated areas. *Sensors* 7, 2519–2538. <https://doi.org/10.3390/s7112519>.
- Thenkabail, P.S., Biradar, C.M., Noojipady, P., Dheeravath, V., Li, Y., Velpuri, M., Gumma, M.K., Gangalakunta, O.R.P., Turrall, 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. <https://doi.org/10.1080/01431160802698919>.
- 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 Sens. (Basel)* 2, 211–261. <https://doi.org/10.3390/rs2010211>.
- Tingting, L., Chuang, L., 2010. Study on extraction of crop information using time-series MODIS data in the Chao Phraya Basin of Thailand. *Adv. Space Res.* 45, 775–784. <https://doi.org/10.1016/j.asr.2009.11.013>.
- Torbick, N., Ledoux, L., Salas, W., Zhao, M., 2016. Regional mapping of plantation extent using multisensor imagery. *Remote Sens. (Basel)* 8, 236. <https://doi.org/10.3390/rs8030236>.
- Uda, S.K., Hein, L., Sumarga, E., 2017. Towards sustainable management of Indonesian tropical peatlands. *Wetl. Ecol. Manag.* 25, 683–701. <https://doi.org/10.1007/s11273-017-9544-0>.
- UN DESA, 2017. World population prospects the 2017 revision key findings and advance tables. *World Popul. Prospect.* 2017, 1–46. <https://doi.org/10.1017/CBO9781107415324.004>.
- USDA NASS, 2014. *2012 Census of Agriculture, 2012 Census of Agriculture United States Summary and State Data*. Washington DC.
- Vuolo, F., Ng, W.-T., Atzberger, C., 2017. Smoothing and gap-filling of high resolution multi-spectral time series: example of Landsat data. *Int. J. Appl. Earth Obs. Geoinf.* 57, 202–213. <https://doi.org/10.1016/j.jag.2016.12.012>.
- Waldner, F., Canto, G.S., Defourny, P., 2015. Automated annual cropland mapping using knowledge-based temporal features. *ISPRS J. Photogramm. Remote Sens.* 110, 1–13. <https://doi.org/10.1016/j.isprsjprs.2015.09.013>.
- Waldner, F., Fritz, S., Di Gregorio, A., Plotnikov, D., Bartalev, S., Kussul, N., Gong, P., Thenkabail, P.S., Hazeu, G., Klein, I., Löw, F., Miettinen, J., Dadhwal, V., Lamarche, C., Bontemps, S., Defourny, P., 2016. A Unified Cropland Layer at 250 M for Global Agriculture Monitoring. *Data*. <https://doi.org/10.3390/data1010003>.
- Wickham, H., Chang, W., 2016. Package ‘ggplot2’. CRAN. <https://doi.org/10.1093/bioinformatics/btr406>.
- World Bank, 2016. *Climate Change Knowledge Portal*. [WWW Document]. URL. (accessed 12.18.17). http://sdwebx.worldbank.org/climateportal/index.cfm?page=country_historical_climate&ThisRegion=Asia&ThisCCode=THA.
- Xiong, J., Thenkabail, P.S., Gumma, M.K., Teluguntla, P., Poehnel, J., Congalton, R.G., Yadav, K., Thau, D., 2017a. Automated cropland mapping of continental Africa using Google Earth Engine cloud computing. *ISPRS J. Photogramm. Remote Sens.* 126, 225–244. <https://doi.org/10.1016/j.isprsjprs.2017.01.019>.
- Xiong, J., Thenkabail, P.S., Tilton, J.C., Gumma, M.K., Teluguntla, P., Oliphant, A.J., Congalton, R.G., Yadav, K., Gorelick, N., 2017b. Nominal 30-m Cropland extent map of continental Africa by integrating pixel-based and object-based algorithms using sentinel-2 and Landsat-8 Data on Google Earth Engine. *Remote Sens. (Basel)* 9, 1065. <https://doi.org/10.3390/rs9101065>.
- Zhou, Y., Xiao, X., Qin, Y., Dong, J., Zhang, G., Kou, W., Jin, C., Wang, J., Li, X., 2016. Mapping paddy rice planting area in rice-wetland coexistent areas through analysis of Landsat 8 OLI and MODIS images. *Int. J. Appl. Earth Obs. Geoinf.* 46, 1–12. <https://doi.org/10.1016/j.jag.2015.11.001>.
- Zhu, Z., Wang, S., Woodcock, C.E., 2015. Improvement and expansion of the Fmask algorithm: cloud, cloud shadow, and snow detection for Landsats 4–7, 8, and Sentinel 2 images. *Remote Sens. Environ.* 159, 269–277. <https://doi.org/10.1016/j.rse.2014.12.014>.