

**Mapping Smallholder Forest Plantations in Andhra Pradesh, India using Multitemporal  
Harmonized Landsat Sentinel-2 S10 Data**

Paige Tatum Williams

Thesis submitted to the faculty of the Virginia Polytechnic Institute and State University in  
partial fulfillment of the requirements for the degree of

Master of Science  
In  
Forestry

Valerie A. Thomas, Chair  
Randolph H. Wynne, Co-Chair  
Ruth DeFries

December 9, 2019  
Blacksburg, VA

Keywords: remote sensing, random forest, NDVI, trees outside forests, machine learning,  
classification

# Mapping Smallholder Forest Plantations in Andhra Pradesh, India using Multitemporal Harmonized Landsat Sentinel-2 S10 Data

Paige Tatum Williams

## ABSTRACT

The objective of this study was to develop a method by which smallholder forest plantations can be mapped accurately in Andhra Pradesh, India using multitemporal (intra- and inter-annual) visible and near-infrared (VNIR) bands from the Sentinel-2 MultiSpectral Instruments (MSIs). Dependency on and scarcity of wood products have driven the deforestation and degradation of natural forests in Southeast Asia. At the same time, forest plantations have been established both within and outside of forests, with the latter (as contiguous blocks) being the focus of this study. The ecosystem services provided by natural forests are different from those of plantations. As such, being able to separate natural forests from plantations is important. Unfortunately, there are constraints to accurately mapping planted forests in Andhra Pradesh (and other similar landscapes in South and Southeast Asia) using remotely sensed data due to the plantations' small size (average 2 hectares), short rotation ages (often 4-7 years for timber species), and spectral similarities to croplands and natural forests. The East and West Godavari districts of Andhra Pradesh were selected as the area for a case study. Cloud-free Harmonized Landsat Sentinel-2 (HLS) S10 data was acquired over six dates, from different seasons, as follows: December 28, 2015; November 22, 2016; November 2, 2017; December 22, 2017; March 1, 2018; and June 15, 2018. Cloud-free satellite data are not available during the monsoon season (July to September) in this coastal region. *In situ* data on forest plantations, provided by collaborators, was supplemented with additional training data representing other land cover subclasses in the region: agriculture, water, aquaculture, mangrove, palm, forest plantation, ground, natural forest, shrub/scrub, sand, and urban, with a total sample size of 2,230. These high-quality samples were then aggregated into three land use classes: non-forest, natural forest, and forest plantations. Image classification used random forests within the Julia Decision Tree package on a thirty-band stack that was comprised of the VNIR bands and NDVI images for all dates. The median classification accuracy from the 5-fold cross validation was 94.3%. Our results, predicated on high quality training data, demonstrate that (mostly smallholder) forest plantations can be separated from natural forests even using only the Sentinel 2 VNIR bands when multitemporal data (across both years and seasons) are used.

# **Mapping Smallholder Forest Plantations in Andhra Pradesh, India using Multitemporal Harmonized Landsat Sentinel-2 S10 Data**

Paige Tatum Williams

## **GENERAL AUDIENCE ABSTRACT**

The objective of this study was to develop a method by which smallholder forest plantations can be mapped accurately in Andhra Pradesh, India using multitemporal (intra- and inter-annual) visible (red, green, blue) and near-infrared (VNIR) bands from the European Space Agency satellite Sentinel-2. Dependency on and scarcity of wood products have driven the deforestation and degradation of natural forests in Southeast Asia. At the same time, forest plantations have been established both within and outside of forests, with the latter (as contiguous blocks) being the focus of this study. The ecosystem services provided by natural forests are different from those of plantations. As such, being able to separate natural forests from plantations is important. Unfortunately, there are constraints to accurately mapping planted forests in Andhra Pradesh (and other similar landscapes in South and Southeast Asia) using remotely sensed data due to the plantations' small size (average 2 hectares), short rotation ages (often 4-7 years for timber species), and spectral (reflectance from satellite imagery) similarities to croplands and natural forests. The East and West Godavari districts of Andhra Pradesh were selected as the area for a case study. Cloud-free Harmonized Landsat Sentinel-2 (HLS) S10 images were acquired over six dates, from different seasons, as follows: December 28, 2015; November 22, 2016; November 2, 2017; December 22, 2017; March 1, 2018; and June 15, 2018. Cloud-free satellite data are not available during the monsoon season (July to September) in this coastal region. In situ data on forest plantations, provided by collaborators, was supplemented with additional training data points (X and Y locations with land cover class) representing other land cover subclasses in the region: agriculture, water, aquaculture, mangrove, palm, forest plantation, ground, natural forest, shrub/scrub, sand, and urban, with a total of 2,230 training points. These high-quality samples were then aggregated into three land use classes: non-forest, natural forest, and forest plantations. A forest cover map in the study area was created from this dataset. The accuracy from the 5-fold cross validation was 94.3%. Our results, predicated on high quality training data, demonstrate that (mostly smallholder) forest plantations can be separated from natural forests even using only the Sentinel 2 VNIR bands when multitemporal data (across both years and seasons) are used.

## Table of Contents

List of Figures .....	v
List of Tables .....	v
1. Introduction.....	1
2. Objective.....	3
3. Study Area .....	3
4. Materials and Methods.....	5
4.1. Model implementation overview .....	5
4.2. Classification scheme.....	6
4.3. Harmonized Landsat Sentinel S10.....	6
4.4. Training and validation data .....	7
4.5. Connecting spectral response to land cover class predictors .....	8
4.6. Separation of vegetation type using NDVI .....	8
4.7. Partitioning and separability of the spectrum.....	10
4.8. Random forests .....	11
4.9. Accuracy assessment .....	12
5. Results.....	12
5.1. Classification map.....	12
5.2. Accuracy assessment .....	16
6. Discussion.....	16
7. Conclusions.....	18
8. References.....	18
Appendix A: Code for Random Forest Classification in Julia.....	24

## List of Figures

<b>Figure 1.</b> Study area: East and West Godavari districts of Andhra Pradesh, India.....	4
<b>Figure 2.</b> Flow diagram of sequential production of this study .....	5
<b>Figure 3.</b> Sample plantation plots shown in different resolutions.....	7
<b>Figure 4.</b> NDVI separation of <i>agriculture</i> and <i>forest plantation</i> example.....	9
<b>Figure 5.</b> NDVI boxplots separation of <i>natural forest</i> and <i>forest plantation</i> .....	9
<b>Figure 6.</b> Feature space image of one date with training and validation data .....	10
<b>Figure 7.</b> Canonical plot of all dates with training and validation data .....	11
<b>Figure 8.</b> Classification output for random forest classification model .....	13
<b>Figure 9.</b> Successful plantation detection zoom-in example.....	14
<b>Figure 10.</b> Training and validation dataset image snapshots .....	15

## List of Tables

<b>Table 1.</b> Study land cover classes with corresponding Bhuvan LULC classification scheme.....	6
<b>Table 2.</b> Training and validation land cover classes.....	8
<b>Table 3.</b> Julia Computing Decision Trees parameter value and description .....	12
<b>Table 4.</b> Average error matrix cross-validation on landcover classes .....	16
<b>Table 5.</b> Summary statistics on accuracy metrics.....	16

## 1. Introduction

Conservation of the world's forests has a renewed importance amid both climate and land use changes, particularly in tropical ecosystems across the world. There is demand for highly accurate spatiotemporal quantifications of global forest cover. However, current global models and many studies fail to distinguish between natural and planted forest types, thus altering the true measure of forest area (Anil, 2011; Hansen et al., 2013; Kayet and Pathak, 2015; Prasad et al., 2015; Sreenivasulu et al., 2013; Puyravaud et al., 2010). Separation of forest types is imperative considering the differences of ecological and socio-economic utility among planted and natural forests (Koskinen et al., 2019).

The critical importance of ecosystem services provided by planted forests will increase in the future due to new opportunities in a globalized market attributed to improvements in wood products and processing technologies. Sustainable and intensively managed planted forests continue to support the growing demand for forest products from timber and wood fiber to oils and fruits (Peterson et al., 2016). In addition to the products extracted from trees, plantations support several external ecosystem services such as clean water, carbon sequestration, regulation of the hydrological cycle, connectivity of habitat fragmentation for biodiversity conservation, and mitigation for deforestation (Kanninen et al., 2010). Dependency and scarcity of wood products has driven the deforestation and degradation of natural forests in Southeast Asia, thus mapping planted forests and natural forests separately will better document the distribution of natural versus anthropogenic systems.

Difficulties in separating natural from managed forests are further exacerbated by a certain degree of definitional limbo within the scientific community as to what, precisely, is considered a forest. This is due to a number of factors, including whether forests are being defined as a land use or a land cover, how society interacts with the forest, and the wide diversity of forest ecosystems around the world. Because of this, the Food and Agriculture Organization Global Forest Resources Assessment of 2000 (FAO FRA) compiled over 650 definitions of forests used in developing countries, and attempted to reduce these definitions into a set of global forest classes that could be applied more consistently (FAO 2001), while still enabling some national modifications where appropriate. This enables comparison of trends in forest cover across nations, and a periodic global accounting of forest cover. As defined by for the FAO FRA 2015, a forest is land spanning over 0.5 ha with tree height above 5 meters and a 10% or more canopy cover, or trees that can meet these thresholds *in situ*. FAO's definition of forest excludes tree stands in an agricultural production system like fruit tree plantations, oil palm plantations, olive orchards, and other agroforestry systems. Their definition of planted forest is "forest predominantly composed of trees established through planting and/or deliberate seeding." In the 2017 Forest Survey of India, the forest class consists of very dense, moderately dense, and open forest including mangrove cover. However, the land use type "land under miscellaneous tree crops and groves" is not considered as part of the recorded forest area and small plantations are considered "trees outside forests" if they fall outside established mapped "greenwash" areas. Given the prevalence of plantations on small land holdings, this definitional exclusion has led to a differential estimation of tree cover in some regions from both the FAO and Forest Survey of India definitions, which has implications for carbon accounting and the monitoring of other ecosystem services in India.

In Southeast Asia, smallholder forest plantations have gained popularity and are replacing degraded or unproductive crop land, offsetting the demand on primary forests (Binkley, 2003;

Okarda et al., 2018; Putra et al., 2019; Puyravaud et al., 2010; Roy et al., 2015; Rudel, 2009; Sharma et al., 2018). Clonal plantations are common in planted forests to genetically improve the growing stock and produce a fast, high yield stock of species such as *Eucalyptus globus* and *Casuarina spp.* (Sharma et al., 2018). Also, palm tree species, notably oil palm and coconut, have rapidly expanded across Southeast Asia due to global market demand (Kannan et al., 2017; Putra et al., 2019). Smallholder farmers, local people in the rural tropics cultivating personal land for subsistence and commercial purposes, are known for their self-initiated forest plantation establishment on plots from one to a few hectares (Pokorny et al., 2010). Paper industries are also seeking available waste and barren land for forest plantation establishment. (Rudel, 2009; Sharma et al., 2018). This land use conversion reduces intensive exploitation of remote primary natural forests.

Multitemporal and multispectral remote sensing data have been widely used for land cover mapping through utilization of relationships between reflectance and vegetation. However, accurately mapping forest plantations in Southeast Asia using remotely-sensed data has been historically constrained by the following: (1) many plantations are small (averaging 2 ha) relative to widely available moderate resolution earth resource satellite data (Lechner et al., 2009), (2) rotation ages for fiber plantations are short (often just 4-7 years) (Sharma et al., 2018), (3) newly established or recently harvested plantations are very difficult to identify correctly (FSI, 2013), and (4) the surrounding cropland area is very variegated in both time and space. Spatial resolution has been one of the main limitations to mapping smallholder forest plantations. Furthermore, while plantations have the potential to be spectrally similar to some agricultural land uses (Griffiths et al., 2019) and natural forest (Behera et al., 2001), they are harvested less frequently than crops, but nonetheless on a regular cycle. A natural forest, in contrast, experiences seasonality but (typically) no harvest. The temporal differences between these two forest types appear tailor-made for interannual multitemporal analysis of remotely sensed data.

Vegetation indices (VIs) derived from remotely sensed data enable separation of vegetated from non-vegetated land cover classes. Spectral reflectance is sensitive to photosynthetic activity in the visible and near infrared bands (Morin et al., 2019). The normalized difference vegetation index (NDVI) is widely used in forest remote sensing because of its association with leaf area and canopy cover, enabling mapping of forests and their condition (le Maire et al., 2011; Nitze et al., 2014; Petitjean and Weber, 2014; Zhu and Liu, 2014; Sheeren et al., 2016; Shilo, 2018). NDVI uses two bands, red and near infrared, in an equation to produce a single value between -1 and 1. The NDVI provides a differencing numerator and a normalizing denominator as shown in equation 1:

$$\frac{NIR - Red}{NIR + Red} \quad (\text{Eq. 1})$$

On satellite imagery the canopy of a mature forest plantation looks visually similar to a natural forest and a young plantation appears similar to many crop types. All vegetative land cover types act very differently across time, and current models fail to differentiate smallholder plantations from natural forest and cropland (Anil, 2011; Hansen et al., 2013; Kayet and Pathak, 2015; Prasad et al., 2015; Reddy et al., 2016b; Sreenivasulu et al., 2013). Different approaches to mapping forest plantations have many tradeoffs considering the wide variety of freely available remotely sensed data and land cover modeling algorithms. MODIS has been commonly used in land cover analysis due to the high frequency of image acquisitions, although a significant

downfall is the spatial resolution (250 m) that does not permit detection of smallholder forest plantations, shifting focus to only large-scale plantations. Remotely sensed data has been used across numerous studies for forest plantation mapping using optical imagery from Landsat (Coleman et al., 1990; Nooni et al., 2014; Kayet and Pathak, 2015; Peterson et al., 2016) and MODIS (le Maire et al. 2011; Miettinen et al., 2012; Jia et al., 2016). There has also been a large body of work in which optical imagery was fused with radar data from ALOS PALSAR (L-band) or Sentinel-1 (C-band; Pin Koh et al., 2011; Tobrick et al., 2016; Koskinen et al., 2019; Poortinga et al., 2019). Peterson et al., (2016) tabulate previous studies mapping forest plantations (mainly oil palm), including their methods, imagery, and accuracy percentages. The two with the highest accuracies (albeit with no focus on smallholders) use a supervised decision tree classifier with 30 m Landsat imagery.

The National Remote Sensing Centre (NRSC) in India produces a periodic land cover classification model for the Forest Survey of India (FSI) using LISS-III data (23.5 m spatial resolution). Their current classification protocol uses satellite imagery from October to December using the green, red, NIR, and SWIR bands. Post monsoonal data is optimal, considering low cloud cover and the post monsoonal flush of leaves which enhances detection of the vegetation types. Proclaimed limitations to forest plantation detection in this assessment include the following: low spatial resolution compared to the average plantation size, non-availability of appropriate seasonal data, mixed classes with forest areas adjacent to cropland, young plantations and trees with less chlorophyll due to low leaf area index and transmittance, and high heterogeneity of tree species (FSI, 2017).

Few to no studies have successfully mapped smallholder forest plantations in South and Southeast Asia. High spatial resolution data are clearly needed. However, while licensed very high spatial resolution data are available from numerous commercial or state entities, only Sentinel-2 VNIR data, at 10 m resolution, have strong potential for smallholder plantation mapping at no cost. Sentinel-2 VNIR data are widely used in land cover and land use change (LCLUC) science for vegetation mapping (Immitzer et al., 2016; Pesaresi et al., 2016; Thanh Noi and Kappas, 2017; Belgiu and Csillik, 2018; Khaliq et al., 2018; Jin et al., 2019), but, as noted above, SWIR bands are commonly used by FSI and other entities to separate plantations from other land uses. However, use of multitemporal data to capture spectral variability across seasons (e.g., Poortinga et al., 2019) and years has the potential to obviate the challenges associated with use of the VNIR data alone.

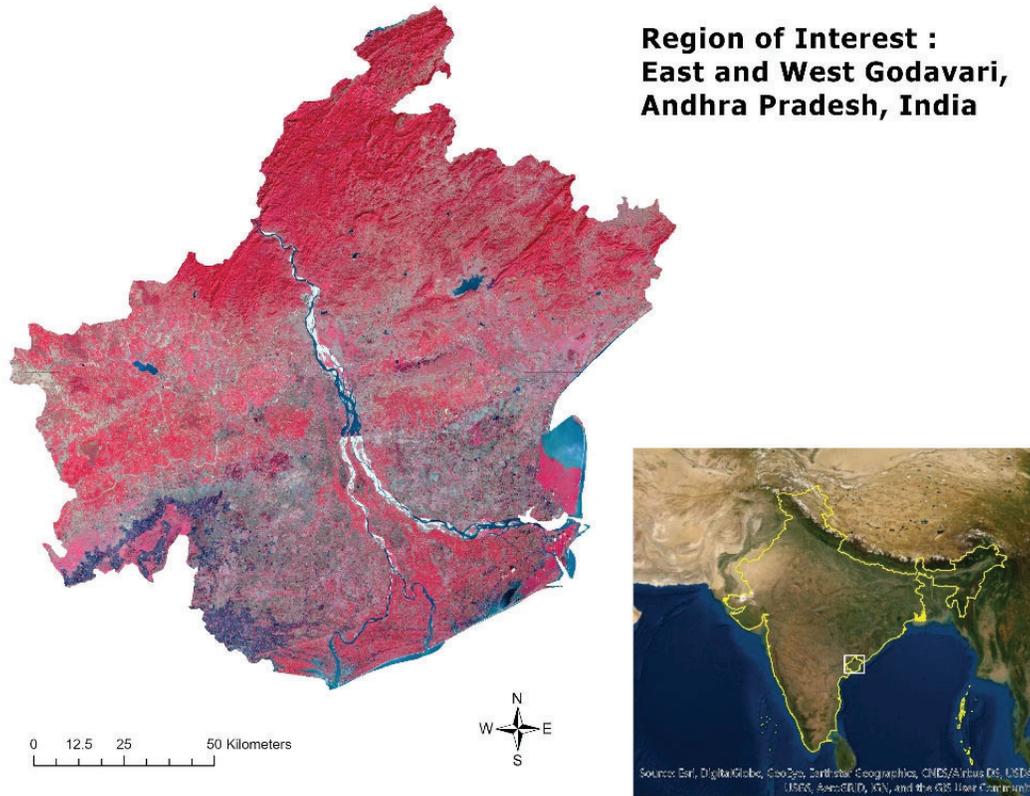
## **2. Objective**

The objective of this study was to develop a method by which smallholder forest plantations can be mapped accurately in Andhra Pradesh, India using multitemporal (intra- and inter-annual) visible and near-infrared (VNIR) bands from Sentinel-2. For the purposes of this study, forest plantations are a combination of FSI's two plantation classes (agriculture plantation and forest plantation).

## **3. Study Area**

In this study we focus on the two districts in Andhra Pradesh, India surrounding the Godavari River: East Godavari and West Godavari (Figure 1). The total area is of both districts combined is 18,501 km<sup>2</sup> and it is located in the southeast region of India between 16°15' N and 18°00' N latitude and 81°00' E 82°20' E longitude. This tropical region experiences three

different seasons: winter (October-February), summer (March-June), and monsoon (July-September). During the monsoon season, these districts receive rainfall from the southwest monsoon from June to September, as well as the northeast monsoon through October and into November (Reddy et al., 2016a). Rainfall exceeds 1,100 mm during the monsoon season, while only 30 mm of rain can be expected to fall between December and March. The annual average temperature is 31.5 °C, with the cooler winter months averaging around 28 °C and the hot, humid summer months reaching 40 °C (Pike, 2018). The northern part of East and West Godavari is home to the discontinuous hills of India's Eastern Ghats.

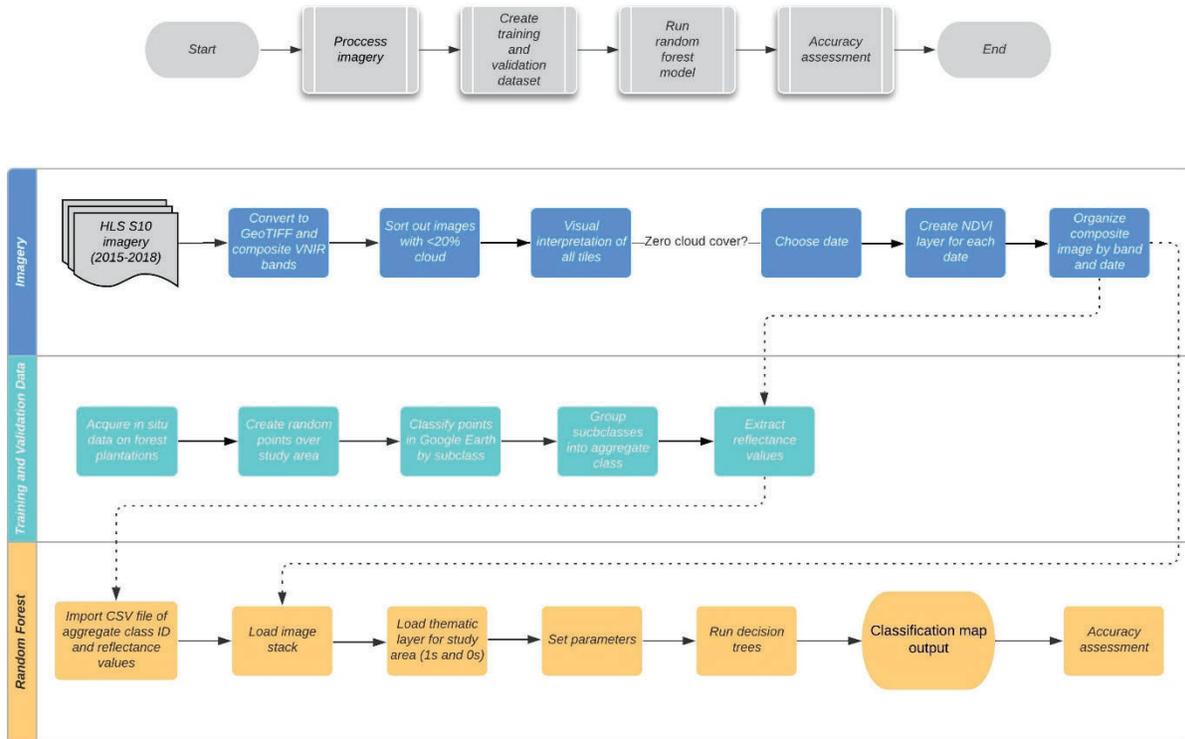


**Figure 1.** Study area: East and West Godavari districts of Andhra Pradesh, India. Harmonized Landsat Sentinel-2 S10 image from December 2015 shown as a standard false color composite on left.

With a population of 49.4 million, Andhra Pradesh is prone to population growth furthering urbanization that is expected to exacerbate deforestation. Nevertheless, the region experienced an increase in recorded forest area for the 2017 assessment due to plantation and conservation activities (FSI, 2017; Anil, 2011). Andhra Pradesh's forest cover area in the state, including forest cover within and outside recorded forest area, is 37,258 km<sup>2</sup> which is 22.9% of the total state area. The dominant forest types in this region include a majority of southern tropical mixed moist deciduous forests, with some patches of semi-evergreen forests (Aditya and Ganesh, 2018). Between the districts, East Godavari has higher total forest cover at 4,726 km<sup>2</sup>. This is due in part to the presence of a natural forest reserve of over 1,000 km<sup>2</sup> in the northern region of the district, known as the Papikondalu National Park (FSI, 2017). At the top of the Eastern Ghats, Papikondalu National Park is known for its densely forested hills, valleys, deep gorges, and streams supplying life to a rich biodiversity of flora and fauna.

## 4. Materials and Methods

This study used a commonly used supervised machine learning method, random forests, to map smallholder forest plantations using remotely sensed data. This machine learning approach includes preparation and processing of the satellite imagery, creation of a training and validation dataset, construction and implementation of the classification model, and an assessment of model accuracy. A flow chart illustrating the sequential production and processing of this land cover classification effort is shown as Figure 2.



**Figure 2.** Flow diagram outlining the sequential production of input data and processing imagery to run a supervised random forest classification model to map forest plantations for the study

### 4.1. Model implementation overview

Target (land use class) and predictor (VNIR reflectances and NDVI for six Sentinel dates) variables were required for training and validation of the random forest classification (Breiman, 1999). The target land use classes were nonforest, natural forest, and plantation; see the next section for the corresponding FSI land use land cover classes. The predictor variables were composited into an image stack of the 30 bands of VNIR and NDVI values from all dates. Training and validation data consisted of 2,230 points. Five-fold cross validation was used for accuracy assessment, in which each fold had 446 samples.

#### 4.2. Classification scheme

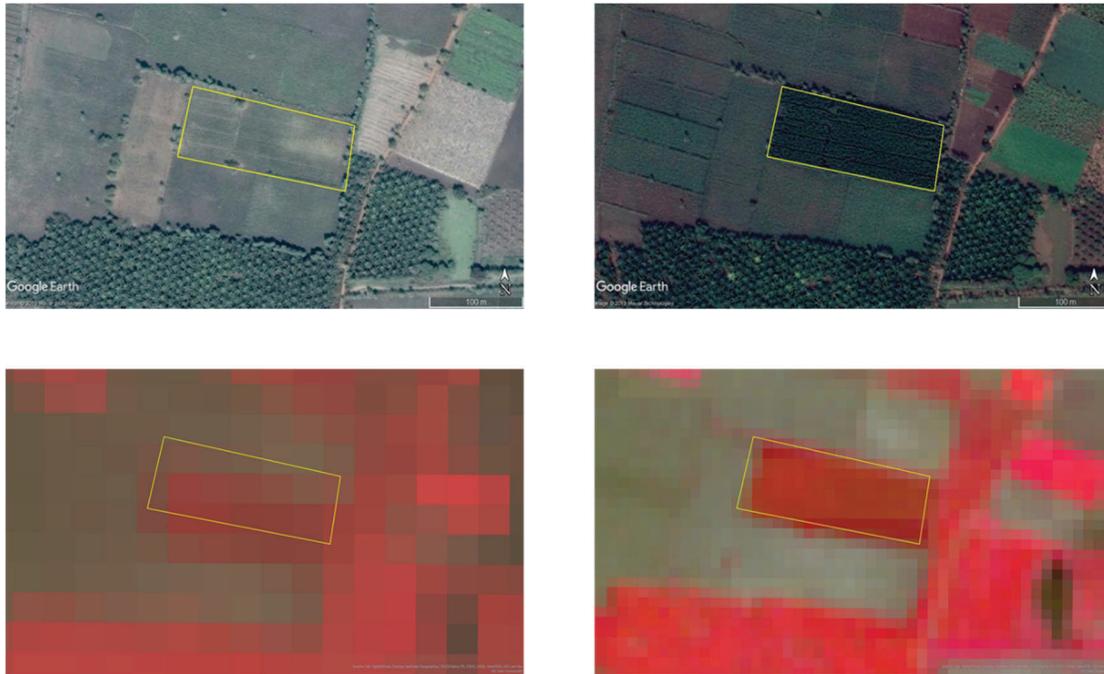
Table 1 show's this study's classification scheme (aggregate classes) cross-walked with the Bhuvan land use and land cover scheme.

*Table 1. Study land cover classes with corresponding FSI land use and land cover classification scheme.*

Study Land Cover Classes	FSI Land Cover Classes
Nonforest	Built up (rural)
	Built up (urban)
	Mining / Industrial
	Crop land
	Aquaculture
	Wastelands
	Waterlogged
	Reservoir / Tanks
	Canal
River / Stream / Drain	
Natural Forest	Forest
	Mangrove
Forest Plantation	Agriculture Plantation
	Forest Plantation

#### 4.3. Harmonized Landsat Sentinel S10

High spatial resolution was a necessity for this study given that the average plantation size is 2 ha. As such, we used the S10 data product from the NASA Harmonized Landsat Sentinel (HLS) program ([hls.gsfc.nasa.gov](https://hls.gsfc.nasa.gov)). The S10 product provides Sentinel-2 MultiSpectral Instrument (MSI) imagery in a UTM grid with BRDF-corrected surface reflectance at full resolutions (10 m, 20 m, 60 m) obtained from L1C products processed by the ESA. The term harmonized signifies the use of a common gridding system (resolution, projection, and spatial extent), radiative transfer algorithm to atmospherically correct to surface reflectance (multiplied by 10,000 and represented as a signed 16-bit integer), nadir view geometry normalized by bidirectional reflectance distribution function (BRDF) estimation, and a spectral bandpass adjustment. Along with the atmospheric correction is a series of cloud metrics integrated into the metadata attributes to estimate percent cloud cover for an image (Claverie et al., 2018). Sentinel-2 MSI (10 m) was used instead of Landsat 8 OLI (30 m) because its finer spatial resolution was preferable for smallholder plantation detection (Figure 3).



**Figure 3.** Sample plantation plot shown in different resolutions. First row: Google Earth images (February 2016, January 2018). Second row: Landsat (30 m) and Sentinel-2 (10 m) from June 2018.

The time at which imagery is acquired plays a key role in land cover classification, considering factors like cloud cover and seasonality of crops (Matton et al., 2015; Morin et al., 2019; Nitze et al., 2014; Zhang et al., 2009). All HLS S10 images covering the study area were acquired from 2015-2018. Our study area covered six of the S10 tiles based on the Sentinel-2 MSI tiling system: QNE, QND, QME, QMD, QPE, and QPD. The data came in the JPG 2000 file type. A MATLAB script was written to convert the files to GeoTIFF format using the image metadata while concomitantly stacking the VNIR (10 m) bands and mosaicking the tiles together. Another MATLAB script sorted through the TIFF files to distinguish images with less than 20% cloud cover using the cloud cover attribute. These images were then visually interpreted in ENVI to select images with zero cloud cover across all tiles. The following dates were chosen: December 28, 2015; November 22, 2016; November 2, 2017; December 22, 2017; March 1, 2018; and June 15, 2018. An NDVI layer was calculated from each of the six HLS images in ENVI using Band Math. NDVI was multiplied by 10,000 and represented as a signed 16-bit integer to correspond with the reflectance representation.

All optical imagery from HLS, plus the six NDVI bands, was combined into a single image stack consisting of 30-bands. Using ERDAS Imagine, the HLS images and NDVI layers were chronologically ordered into a 30-band image stack; five bands across all six dates. Each of the layers in the 30-band image stack were named by band and date (year and day of year). The final image stack was then clipped to a shapefile of the study area.

#### 4.4. Training and validation data

Construction of the training dataset required multiple, extensive random point assessments to achieve complete and clear representation of land cover classes in this region of

study, resulting in a final training dataset of 2,230 points aggregated into three main land cover classes: nonforest, natural forest, and plantation (Table 2). In situ data was collected by our team during several weeks of field work in December 2018. Collaborators from International Paper assisted in identification of forest plantation types in this region. Points were created in ArcMap using the Random Points tool and loaded into Google Earth using available high-resolution imagery (sub-meter resolution from Digital Globe) for visual interpretation.

Our classification scheme required consideration of the multitemporal nature of the image data and phenological and spectral variability within a main class, therefore photo-interpretation of the Digital Globe imagery required the following rules: each point had to be consistently the same sub-class through time (2015-2018), a 10-meter buffer surrounding the point avoided edge pixels, and each point was not mixed with any other subclass. The resulting land cover subclasses include agriculture, aquaculture, ground, sand, urban, shrub/scrub, water, natural forest, mangrove, palm plantations, and forest plantations.

*Table 2. In situ data on forest plantations provided by collaborators was supplemented with additional training data points and aggregated into 3 classes: nonforest, natural forest, and plantation.*

Land Cover Type	Number of Points	Aggregate Class
Agriculture	555	Nonforest <i>n</i> = 1,467
Aquaculture	153	
Ground	81	
Sand	110	
Urban	119	
Shrub/scrub	224	
Water	225	
Natural Forest	241	Natural Forest <i>n</i> = 299
Mangrove	58	
Forest Plantation	253	Plantation <i>n</i> = 464
Palm Plantation	211	
Total: 2,230 points		

#### 4.5. Connecting spectral response to land cover class predictors

An R script was used to extract the 30 band / NDVI values for each sample point, resulting in a comma-separated values (CSV) file. Each row represented one sample point, and contained the aggregate (target) class, subclass, X location, Y location, and the 30 column reflectance / NDVI vector.

#### 4.6. Separation of vegetation type using NDVI

Given the similarities in vegetation spectral signatures, the spectral responses from different crop types and natural forest can be confused with planted forests (Morin et al., 2019; Nitze et al., 2015). NDVI values from a single agriculture and forest plantation (*Casuarina spp.*) training and validation point were graphed across the HLS dates used in the random forest model

(Figure 4). Another analysis using NDVI was performed on all training points included in the two forest type classes: natural forest (mangrove and natural forest) and plantation (fiber and palm plantation). A box plot was implemented to assess the distribution of NDVI values across HLS dates within the two forest type classes (Figure 5).

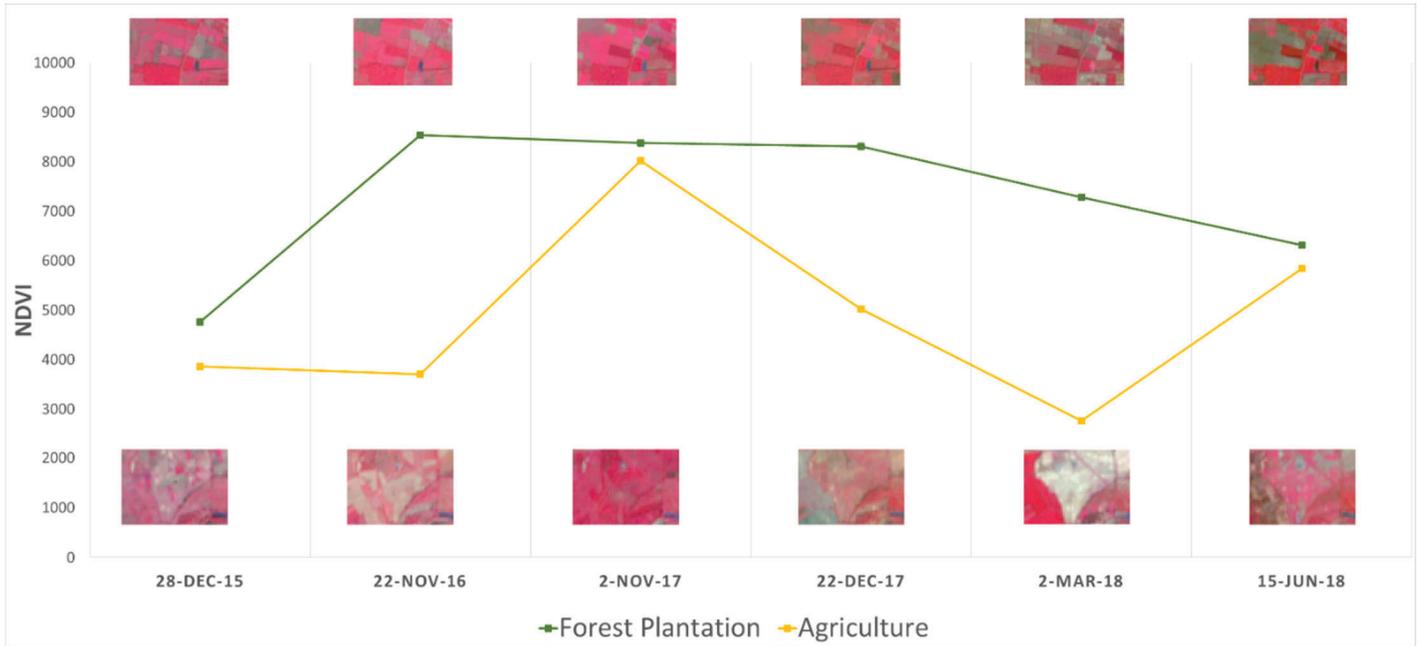


Figure 4. Separation of a forest plantation example (green) from an agriculture field example (yellow) with NDVI response values across HLS dates.

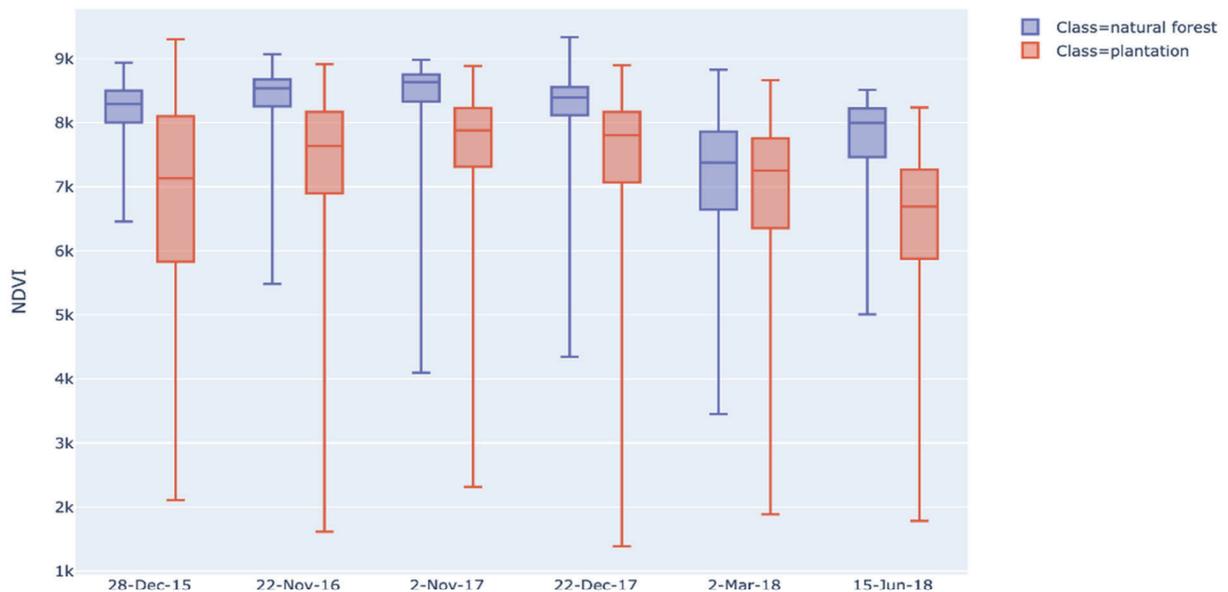
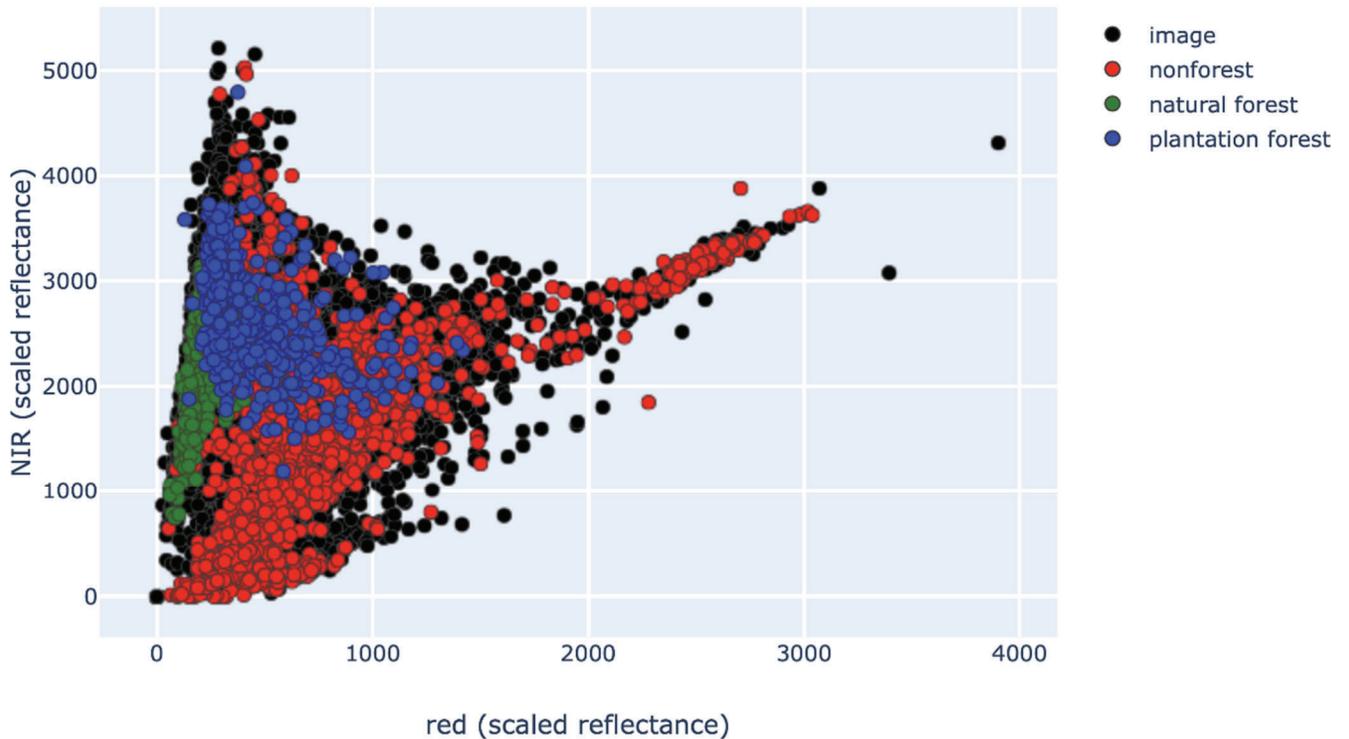


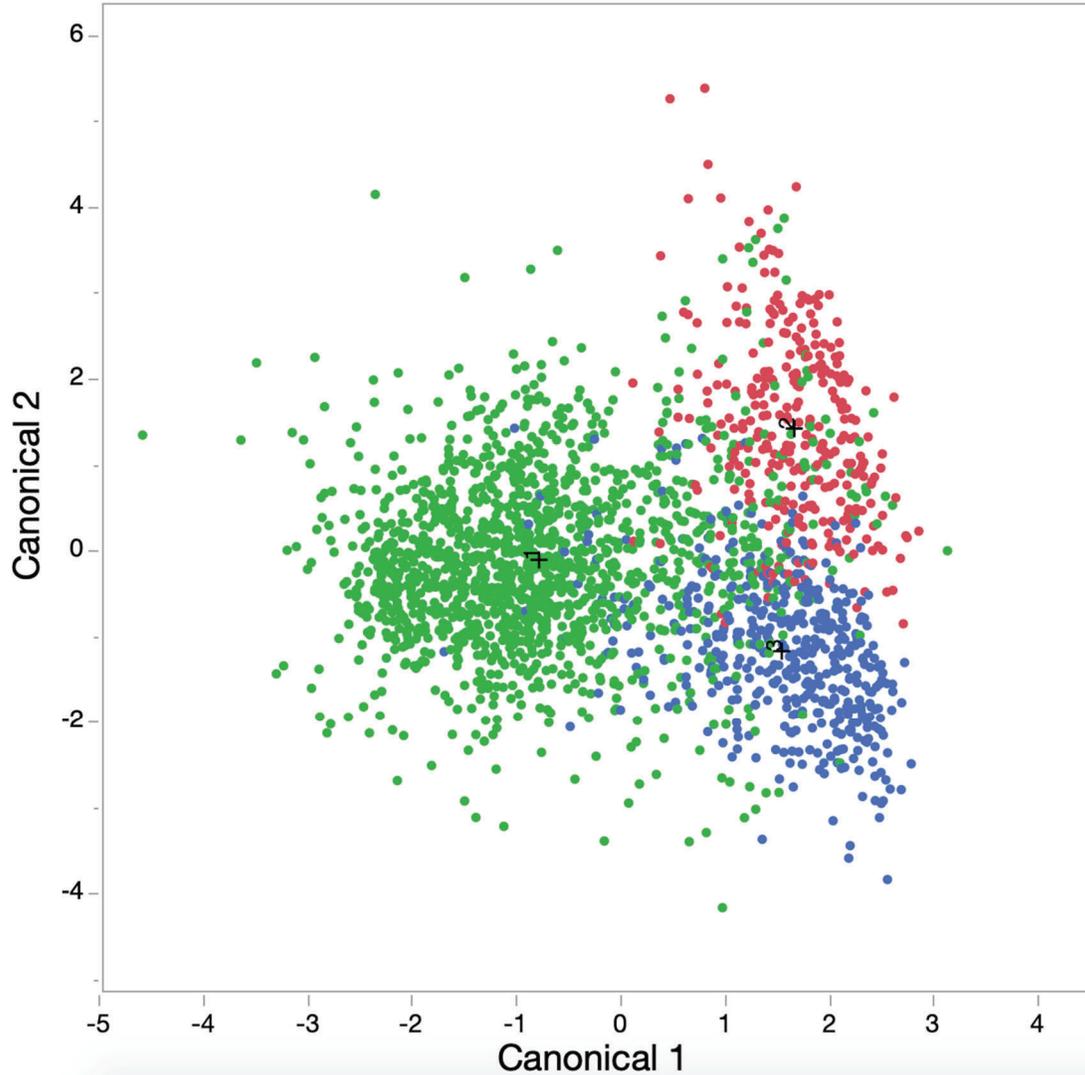
Figure 5. Spectral separation of natural forest (blue) and plantation (red) as boxplots of minimum, first quartile, median, third quartile, and maximum of NDVI response values across HLS dates (n = 2,230). Note that the dates maximizing separability of natural forest from plantation forest are in most cases different from the dates maximizing separability of forest from agriculture (Figure 4).

#### 4.7. Partitioning and separability of the spectrum

For remote sensing land cover classification, it is good practice to optimize partitioning and separability among predictor and response variables (Campbell and Wynne, 2011). For this study, our predictor variables are the three main land cover classes and the response variables are the spectral responses in the visible and near-infrared bands along with NDVI across all six dates. A feature space image comparing the reflectance responses of the red and NIR bands by land cover class from the first date in 2015 is shown as Figure 6, indicating very good to excellent partitioning. Presence of slight class confusion for the forest plantation class is resolved with use of seasonal and interannual multitemporal data as shown in the canonical plot using all 30 VNIR and NDVI bands (Figure 7).



**Figure 6.** Feature space image with red reflectances on the X-axis and NIR reflectances ( $n = 2,230$ ) on the Y-axis showing partitioning of the feature space. HLS pixel values (same two bands systematically sampled from the 28-Dec-15 mosaic with  $n = 34,977$ ) shown in black. Training and validation dataset shown in different colors by land cover class: Nonforest is red, natural forest green, and plantations blue. Note that the feature space is well-partitioned, but this single date does not afford sufficient class separation.



**Figure 7.** Canonical plot of linear combinations of the covariates using all 30 VNIR and NDVI bands showing good class separation ( $n = 2,230$ ). Canonical weights (not shown) were approximately equal among dates. Nonforest is red, natural forest green, and plantations blue.

#### 4.8. Random forests

Multiple machine learning algorithms, including random forests, CART, and SVM, were tested on the dataset. A random forest classifier proved optimal for this large, variegated area. The Julia programming language (version 1.3.0) was chosen for this analysis due to its efficiency and robust memory management. The DecisionTree.jl (version 0.10.0; <https://github.com/bensadeghi/DecisionTree.jl>) package was used to implement random forests. The classification model used in this analysis includes parameters such as pre-pruning (max depth, min leaf size), post pruning (pessimistic pruning), multi-threaded bagging (random forests), adaptive boosting (decision stumps), and cross validation (n-fold). A random forest with 50 trees was selected after an iterative parameter optimization. Table 3 presents the parameters and descriptions used for our model assessment

*Table 3. Julia DecisionTree.jl random forest classifier parameter values and descriptions.*

Parameter	Value	Description
n_folds	5	Number of cross validation subsamples
n_subfeatures	-1	Number of features to consider at random per split (-1 := $\sqrt{n\_features}$ )
n_trees	50	Number of individual decision trees
partial_sampling	.7	Proportion of samples per tree
max_depth	-1	Maximum depth of the decision trees, -1 := no maximum
min_samples_leaf	10	Minimum number of samples each leaf needs to have
min_samples_split	5	Minimum number of samples needed for a split
min_purity_increase	0.0	Minimum purity needed for a split (used for post-pruning)

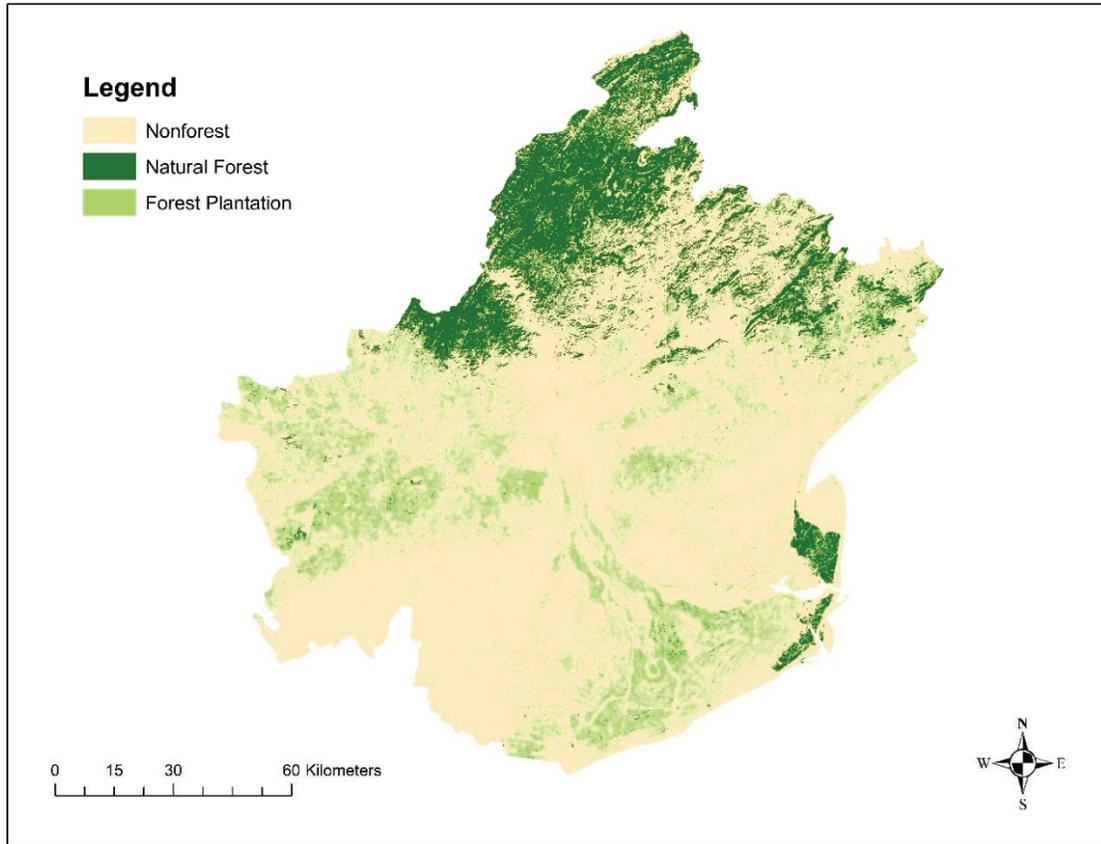
#### 4.9. Accuracy assessment

Model accuracy was estimated from the training and validation dataset using a 5-fold cross-validation, with 446 samples per fold. The error matrix and resulting summary statistics (overall accuracy, kappa, class-specific user's and producer's accuracies) were calculated using standard techniques (Campbell and Wynne, 2011).

## 5. Results

### 5.1. Classification map

The supervised random forest classifier using the Julia DecisionTree.jl package produced a classification map with 10 m resolution over East and West Godavari separated into 3 land cover classes: nonforest (tan), natural forest (dark green), and forest plantation (light green) (Figure 8). The nonforest class includes the majority of the land cover classes present in this region and accounted for 74.5% of total area. Natural forest, including conserved forest in the north and mangroves along the coast, is estimated at 14.5% of total area. The target class, forest plantation, includes palm and other tree plantations and amounts to 11% of total area. Model performance was visually assessed at a fine scale using HLS images, classification output, and a high-resolution base map in ArcGIS Pro by zooming into areas with known land use and land cover. Figure 9, for example, shows the result of this process for fiber plantations in East Godavari. Figure 10 shows the classification output, high-resolution imagery from Google Earth, and HLS images for all land cover classes.



**Figure 8.** Classification map with 10 m resolution over East and West Godavari separated into 3 land cover classes: nonforest (tan), natural forest (dark green), and plantation (light green).

Harmonized Landsat Sentinel-2 S10 Images



Julia Decision Trees Classification Output  
\*Forest plantation in light green



Digital Globe Imagery from ArcGIS Pro Basemap

**Figure 9.** Plantation detection zoom-in example. Top row: HLS imagery across dates utilized. Bottom row: Plantations in light green classified by model (left), and shown on Digital Globe imagery from ArcGIS Pro base map (right). Two known plantation locations are shown as cyan points.

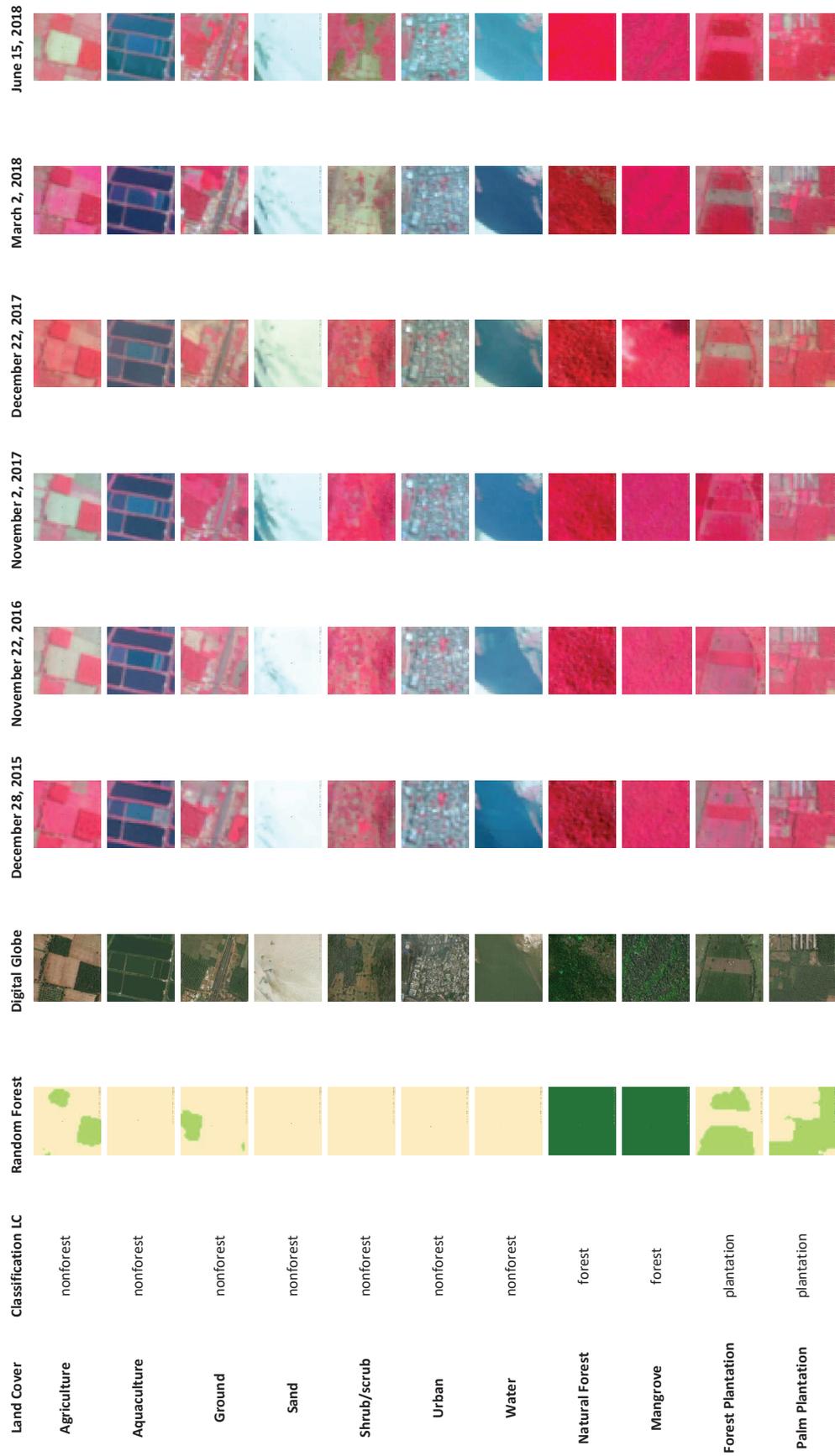


Figure 10. Training and validation dataset with supplemental model classification output, Digital Globe imagery, and HLS images by land cover class.

## 5.2. Accuracy assessment

The validation results are shown using a confusion matrix (Table 5), and accuracy summary statistics (Table 6) from the 5-fold cross-validation, 446 samples per fold. As shown in figures 4 through 7, all utilized dates and bands were important, and iterative, selective elimination of any one date or band produced an evident decrease in model performance. Average overall accuracy across the five folds was 94.3%. The target class, plantations, was successfully classified, but was slightly confused with nonforest. The nonforest class had the highest class-specific accuracies, presumably due to its spectral dissimilarity from forest in the aggregate (excluding agriculture) and its preponderance (65.8% of sample points) in the random (but therefore unbalanced) sample.

*Table 4. Average error matrix from the 5-fold cross-validation, n = 446 samples/fold.*

	<b>Nonforest</b>	<b>Natural Forest</b>	<b>Forest Plantation</b>
<b>Nonforest</b>	282.4	7.4	3.6
<b>Natural Forest</b>	3.4	54.6	1.8
<b>Forest Plantation</b>	7.2	.4	85.2

*Table 5. Accuracy summary statistics calculated from average error matrix (Table 5).*

<b>LC Class</b>	<b>User's Accuracy</b>	<b>Producer's Accuracy</b>	<b>Overall Accuracy</b>	<b>Kappa</b>
Nonforest	96	96.1		
Natural Forest	92	86.3	94.3	88.7
Forest Plantation	90.1	94		

## 6. Discussion

Using both intra- (Jia et al., 2016) and interannual (Poortinga et al., 2019) temporal variation to separate otherwise similar spectral signatures was the cornerstone of this successful classification (see also, e.g., the differences in class separability between Figures 3 and 4). Figure 10 captures the visual spectral variation in false-color HLS image snapshots of different land cover classes compared to the ground reference and model classification. Even a given vegetation type can have different temporal and spectral responses due to differences in local land management, genetic features, site conditions, and many other environmental factors. As such, sampling such that the spectro-temporal feature space is well-partitioned is vital. The use of temporal information enables differentiation of vegetative types using differences in seasonal cycles and vegetation phenology (Griffiths et al., 2019; Zhang et al., 2009).

Capturing the variability of vegetation phenology and seasonal growing characteristics is essential when modeling the separation of cropland and planted forest types (Griffiths et al., 2019; le Maire et al., 2011; Nitze et al., 2014; Zhang et al 2009; Zhu and Liu 2014). Clear sky

observations during the monsoon season are rare to non-existent. As such, the dates used in this analysis include the prominent winter months, where vegetation is at its peak in this region due to water availability, and the summer months to capture vegetation prior to the rainy season when it may be dry or unhealthy. Use of winter and summer dates optimizes separability of vegetation types because the phenology is more stable during these seasons (Morin et al., 2019; Behera et al., 2001).

The model proved successful in separating forest plantations from agriculture in part using (indirectly) the harvest cycles for different crop types. Figure 4 shows harvest and regeneration for an agriculture point, while the forest plantation point grows over time and levels off in the dryer season (summer) when the trees may not be at peak vigor. In this figure the NDVI values for the two planted types do not converge. This specific case was corroborated via preliminary analyses using Sentinel-2 MSI.

Across years and time natural forests have, in general, higher NDVIs than forest plantations (Figure 5). NDVI variability is also greater in forest plantations for all dates except March 2018. The wide variability of plantation NDVIs is likely due to the different types and ages of stands within the plantation class (see the top row of figure 9 for an example of the change in appearance of plantations from establishment to maturity). However, even given this variability, it is clear from Figure 5 that the plantation and natural forest classes are generally separable using NDVI alone.

At the study design phase, we tested imagery from the Landsat 8 Operational Land Imager (30 m) and from the 20-m Sentinel-2 MSI bands (SWIR and red edge). However (see, e.g., Figure 3), neither of these sensors had sufficient resolution to detect smallholder forest plantations as trees outside forests. Even the inclusion of the SWIR bands (both sensors) and red-edge bands (Sentinel-2 MSI) could not compensate for the decreased spatial resolution. Keep in mind, however, that this preliminary analysis was focused on just the identification of plantations without attempting greater categorical specificity (such as species or other taxonomic groupings). Discrimination of tree species in the tropics has been shown to improve using the SWIR (Ferreira et al., 2015).

Forest expansion occurs from two main causes: forest plantation establishment or the spontaneous reforestation of abandoned land (Mather, 2007). For LCLUC science, defining what type of forest is expanding will be vital for ecological and economic modeling. As such, our study focused on three main land cover classes in a hierarchical sampling design: nonforest, natural forest, and forest plantation. This now vetted approach to forest plantation detection can be further utilized in subsequent efforts that map natural vs. planted forests in the region.

Accurate mapping of trees outside forests is important both economically and scientifically. Expansion of forest area has been identified as a possible natural climate solution (Griscom et al., 2017), and accurate carbon accounting will require quantification of trees outside forests as well as those in greenwash areas. Plantation establishment and forest degradation both affect radiative forcing through changes in albedo and biosphere-atmosphere gas exchange. While not a focus of this study, improved monitoring of conditions of native forests can assist in estimation of biodiversity richness and habitat fragmentation (Roy et al., 2013).

High frequency of temporal coverage and high spatial resolution are both imperative for quantifying different forest types across a heterogeneous landscape, where natural forests and plantations are woven in and around each other (Roy et al., 2015). Sentinel-2 MSI data proved sufficient to the task for block plantations in this instance. However, there are other realizations

of trees outside forests, namely windbreaks, scattered trees, and linear plantations (Rawat et al., 2003) that will likely require higher resolution imagery for accurate quantification.

Finally, while we were successful in using supervised machine learning via the commonly utilized random forests algorithm, deep learning is also gaining popularity in remote sensing science. It has strong potential for mapping at this and, in particular, increased levels of categorical specificity (Ienco, 2017).

## 7. Conclusions

Intra- and interannual VNIR reflectance data from Sentinel 2 MSI, coupled with high quality training data that capture spectro-temporal variability, enable fine-scale forest plantation detection in Andhra Pradesh using a common machine learning approach. The spatial resolution and radiometric quality of the Sentinel 2 data, coupled with their availability at no-cost, make them particularly suitable to mapping trees outside forests. Quantifying the ecosystem services provided by smallholder plantation forests in South and Southeast Asia will require regular, accurate mapping to capture both status and change. These future efforts, whether by state or non-state actors, will be engendered by building on the lessons learned from this case study in Andhra Pradesh.

## 8. References

- Aditya, V. & Ganesh, T. (2018). Deciphering forest change: Linking satellite-based forest cover change and community perceptions in a threatened landscape in India. *Ambio*, 48(7), 790–800. <https://doi.org/10.1007/s13280-018-1108-x>
- Anil, N. C., Sankar, G. J., Rao, M. J., & Sailaja, U. (2011). Studies on land use/land cover and change detection from parts of South West Godavari District, A. P. using remote sensing and GIS techniques. *Journal of the Indian Geophysical Union*, 15(4), 187–194.
- Behera, M. D., Kushwaha, S. P. S., & Roy, P. S. (2001). Forest vegetation characterization and mapping using IRS-1C satellite images in Eastern Himalayan Region. *Geocarto International*, 16(3), 53–62. <https://doi.org/10.1080/10106040108542204>
- Belgiu, M., & Csillik, O. (2018). Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. *Remote Sensing of Environment*, 204, 509–523. <https://doi.org/10.1016/j.rse.2017.10.005>
- Binkley, C. S. (2003). Forestry in the long sweep of history. In L. Teeter, B. Cashore, & D. Zhang (Eds.), *Forest policy for private forestry: global and regional challenges* (pp. 1–8). Wallingford, Oxon, United Kingdom: CABI Publishing.
- Breiman, L. (1999). Random Forests--Random Features (Department of Statistics Technical Reports) (pp. 1–29). University of California, Berkeley.

Campbell, J. B., & Wynne, R. H. (2011). *Introduction to Remote Sensing* (5th ed.). New York: Guilford.

Claverie, M., Ju, J., Masek, J. G., Dungan, J. L., Vermote, E. F., Roger, J. C., ... Justice, C. (2018). The Harmonized Landsat and Sentinel-2 surface reflectance data set. *Remote Sensing of Environment*, 219, 145–161. <https://doi.org/10.1016/j.rse.2018.09.002>

Coleman, T. L., Gudapati, L., & Derrington, J. (1990). Monitoring forest plantations using Landsat Thematic Mapper data. *Remote Sensing of Environment*, 33, 211-221. [https://doi.org/10.1016/0034-4257\(90\)90032-H](https://doi.org/10.1016/0034-4257(90)90032-H)

Ferreira, M. P., Zortea, M., Zanotta, D. C., Féret, J. B., Shimabukuro, Y. E., & Souza Filho, C. R. (2015). On the use of shortwave infrared for tree species discrimination in tropical semideciduous forest. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XL-3/W3, 473–476. <https://doi.org/10.5194/isprsarchives-XL-3-W3-473-2015>

FAO. (2001). *State of the world's forests 2001. Part II. Key issues in the forest sector today*. Rome: Food and Agriculture Organization of the United Nations. [http://www.fao.org/3/y0900e/y0900e05.htm#P0\\_0](http://www.fao.org/3/y0900e/y0900e05.htm#P0_0)

Forest Survey of India. (2013). *India State of Forest Report*.

Forest Survey of India. (2017). *India State of Forest Report*.

FRA 2015. (2015). *Forest Resources Assessment Working Paper*.

Griffiths, P., Nendel, C., & Hostert, P. (2019). Intra-annual reflectance composites from Sentinel-2 and Landsat for national-scale crop and land cover mapping. *Remote Sensing of Environment*, 220, 135–151. <https://doi.org/10.1016/j.rse.2018.10.031>

Griscom, B. W., Adams, J., Ellis, P. W., Houghton, R. A., Lomax, G., Miteva, D. A., ... Fargione, J. (2017). Natural climate solutions. *Proceedings of the National Academy of Sciences of the United States of America*, 114(44), 11645–11650. <https://doi.org/10.1073/pnas.1710465114>

Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., ... Townshend, J. R. G. (2013). High-resolution global maps of 21st-century forest cover change. *Science*, 342(6160), 850–853. <https://doi.org/10.1126/science.1244693>

Ienco, Di., Gaetano, R., Dupaquier, C., & Maurel, P. (2017). Land cover classification via multitemporal spatial data by deep recurrent neural networks. *IEEE Geoscience and Remote Sensing Letters*, 14(10), 1685–1689. <https://doi.org/10.1109/LGRS.2017.2728698>

- Immitzer, M., Vuolo, F., & Atzberger, C. (2016). First experience with Sentinel-2 data for crop and tree species classifications in Central Europe. *Remote Sensing*, 8(3), 166. <https://doi.org/10.3390/rs8030166>
- Jia, X., Khandelwal, A., Gerber, J., Carlson, K., West, P., & Kumar, V. (2016). Learning large-scale plantation mapping from imperfect annotators. *2016 IEEE International Conference on Big Data (Big Data)*, 1192–1201. <https://doi.org/10.1109/BigData.2016.7840723>
- Jin, Z., Azzari, G., You, C., Di Tommaso, S., Aston, S., Burke, M., & Lobell, D. B. (2019). Smallholder maize area and yield mapping at national scales with Google Earth Engine. *Remote Sensing of Environment*, 228, 115-128. <https://doi.org/10.1016/j.rse.2019.04.016>
- Kannan, B., Ragonath, K. P., Kumaraperumal, R., Jagadeeswaran, R., & Krishnan, R. (2017). Mapping of coconut growing areas in Tamil Nadu, India using remote sensing and GIS. *Journal of Applied and Natural Science*, 9(2), 771–773. <https://doi.org/10.31018/jans.v9i2.1272>
- Kanninen, M. (2010). Plantation forests: global perspectives. In J. Bauhus, P. van der Meer, & M. Kanninen (Eds.), *Ecosystem goods and services from plantation forests* (pp. 1-15). London: Routledge. <https://doi.org/10.4324/9781849776417>
- Kayet, N., & Pathak, K. (2015). Remote sensing and GIS based land use / land cover change detection. *International Research Journal of Earth Sciences*, 3(10), 1–6.
- Khaliq, A., Peroni, L., & Chiaberge, M. (2018). Land cover and crop classification using multitemporal Sentinel-2 images based on crops phenological cycle. *EESMS 2018 - Environmental, Energy, and Structural Monitoring Systems, Proceedings*. <https://doi.org/10.1109/EESMS.2018.8405830>
- Koskinen, J., Leinonen, U., Vollrath, A., Ortmann, A., Lindquist, E., d'Annunzio, R., ... Käyhkö, N. (2019). Participatory mapping of forest plantations with openforis and Google Earth Engine. *ISPRS Journal of Photogrammetry and Remote Sensing*, 148, 63-74. <https://doi.org/10.1016/j.isprsjprs.2018.12.011>
- le Maire, G., Marsden, C., Nouvellon, Y., Grinand, C., Hakamada, R., Stape, J. L., & Laclau, J. P. (2011). MODIS NDVI time-series allow the monitoring of Eucalyptus plantation biomass. *Remote Sensing of Environment*, 115(10), 2613-2625. <https://doi.org/10.1016/j.rse.2011.05.017>
- Lechner, A. M., Stein, A., Jones, S. D., & Ferwerda, J. G. (2009). Remote sensing of small and linear features: quantifying the effects of patch size and length, grid position and detectability on land cover mapping. *Remote Sensing of Environment*, 113(10), 2194–2204. <https://doi.org/10.1016/j.rse.2009.06.002>
- Mather, A. S. (2007). Recent Asian forest transitions in relation to forest transition theory. *International Forestry Review*, 9(1), 491–502. <https://doi.org/10.1505/ifor.9.1.491>

Matton, N., Management, G. I., Sepulcre-canto, G., Waldner, F., Scientific, T. C., Morin, D., ... La, S. De. (2015). An automated method for annual cropland mapping along the season for various globally-distributed agrosystems using high spatial and temporal resolution time series. *Remote Sensing*, 7 (October), 13208–13232. <https://doi.org/10.3390/rs71013208>

Miettinen, J., Shi, C., Tan, W. J., & Liew, S. C. (2012). 2010 land cover map of insular Southeast Asia in 250-m spatial resolution. *Remote Sensing Letters*, 3(1), 11–20. <https://doi.org/10.1080/01431161.2010.526971>

Morin, D., Planells, M., Guyon, D., Villard, L., Bouvet, A., Thevenon, H., & Dejoux, J. (2019). Estimation and mapping of forest structure parameters from open access satellite images: development of a generic method with a study case on coniferous plantation. *Remote Sensing*, 11 (1275).

Nitze, I., Barrett, B., & Cawkwell, F. (2014). Temporal optimisation of image acquisition for land cover classification with random forest and MODIS time-series. *International Journal of Applied Earth Observation and Geoinformation*, 34, 136-146. <https://doi.org/10.1016/j.jag.2014.08.001>

Nooni, I. K., Duker, A. A., Van Duren, I., Addae-Wireko, L., & Osei Jnr, E. M. (2014). Support vector machine to map oil palm in a heterogeneous environment. *International Journal of Remote Sensing*, 35(13), 4778–4794. <https://doi.org/10.1080/01431161.2014.930201>

Okarda, B., Carolita, I., Kartika, T., & Komarudin, H. (2018). Mapping of smallholder oil palm plantation and development of a growth model. *IOP Conference Series: Earth and Environmental Science*, 169, 012074. <https://doi.org/10.1088/1755-1315/169/1/012074>

Pesaresi, M., Corbane, C., Julea, A., Florczyk, A., Syrris, V., & Soille, P. (2016). Assessment of the added-value of Sentinel-2 for detecting built-up areas. *Remote Sensing*, 8(4), 299. <https://doi.org/10.3390/rs8040299>

Petersen, R., Aksenov, D., Esipova, E., Goldman, E., Harris, N., Kurakina, I., ... Shevade, V. (2016). *Mapping tree plantations with multispectral imagery: Preliminary results for seven tropical countries*. Washington: World Resources Institute.

Petitjean, F., & Weber, J. (2014). Efficient satellite image time series analysis under time warping. *IEEE Geoscience and Remote Sensing Letters*, 11(6), 1143-1147. <https://doi.org/10.1109/LGRS.2013.2288358>

Pike, J. (2018). Andhra Pradesh - Climate. Retrieved from <https://www.globalsecurity.org/military/world/india/andhra-pradesh-climate.htm>.

Pin Koh, L., Miettinen, J., Liew, S. C., & Ghazoul, J. (2011). Remotely sensed evidence of tropical peatland conversion to oil palm. *Proceedings of the National Academy of Sciences of the United States of America*, 108(12), 5127–5132. <https://doi.org/10.1073/pnas.1018776108>

- Pokorny, B., Hoch, L., & Maturana, J. (2010). Smallholder plantations in the tropics – Local people between outgrower schemes and reforestation programmes. In J. Bauhus, P. van der Meer, & M. Kanninen (Eds.), *Ecosystem goods and services from plantation forests* (pp. 140–170). London: Routledge. <https://doi.org/10.4324/9781849776417>
- Poortinga, A., Tenneson, K., Shapiro, A., Nquyen, Q., & San, K. (2019). Mapping plantations in Myanmar by fusing Landsat-8, Sentinel-2 and Sentinel-1 data along with systematic error quantification. *Remote Sensing*, *11*(7), 831. <https://doi.org/10.3390/rs11070831>
- Prasad, P. R. C., & Lakshmi, P. M. (2015). Estimation of biomass density and carbon storage in the forests of Andhra Pradesh, India, with emphasis on their deforestation and degradation conditions. *European Journal of Ecology*, *1*(1), 52–63. <https://doi.org/10.1515/eje-2015-0007>
- Putra, B. T. W., Soni, P., Marhaenanto, B., Pujiyanto, Sisbudi Harsono, S., & Fountas, S. (2019). Using information from images for plantation monitoring: A review of solutions for smallholders. *Information Processing in Agriculture*. <https://doi.org/10.1016/j.inpa.2019.04.005>
- Puyravaud, J. P., Davidar, P., & Laurance, W. F. (2010). Cryptic destruction of India's native forests. *Conservation Letters*, *3*(6), 390–394. <https://doi.org/10.1111/j.1755-263X.2010.00141.x>
- Rawat, J. K., Dasgupta, S., Kumar, R., Kumar, A., & Chauhan, K. V. S. (2003). *Training Manual on Inventory of Trees Outside Forests* (Information and Analysis for Sustainable Forest Management: Linking National and International Efforts in South and Southeast Asia) (pp. 1–39). Bangkok: FAO.
- Reddy, C. S., Jha, C. S., Dadhwal, V. K., Krishna, P. H., Pasha, S. V., Satish, K. V., ... Diwakar, P. G. (2016a). Quantification and monitoring of deforestation in India. *Biodiversity and Conservation*, *25*(1), 93–116. <https://doi.org/10.1007/s10531-015-1033-2>
- Reddy, C. S., Manaswini, G., Satish, K. V., Singh, S., Jha, C. S., & Dadhwal, V. K. (2016b). Conservation priorities of forest ecosystems: Evaluation of deforestation and degradation hotspots using geospatial techniques. *Ecological Engineering*, *91*, 333–342. <https://doi.org/10.1016/j.ecoleng.2016.03.007>
- Roy, P. S., Murthy, M. S. R., Roy, A., Kushwaha, S. P. S., Singh, S., Jha, C. S., ... Gupta, S. (2013). Forest fragmentation in India. *Current Science*, *105*(6), 774–780.
- Roy, P. S., Behera, M. D., Murthy, M. S. R., Roy, A., Singh, S., Kushwaha, S. P. S., ... Ramachandran, R. M. (2015). New vegetation type map of India prepared using satellite remote sensing: Comparison with global vegetation maps and utilities. *International Journal of Applied Earth Observation and Geoinformation*, *39*, 142–159. <https://doi.org/10.1016/j.jag.2015.03.003>
- Rudel, T. K. (2009). Tree farms: Driving forces and regional patterns in the global expansion of forest plantations. *Land Use Policy*, *26*(3), 545–550. <https://doi.org/10.1016/j.landusepol.2008.08.003>

Sharma, S., Arya, I., Ewari, S., Arya, S., & Yadava, M. (2018). Clonal plantations play a key role to increase agroforestry production enriching farm communities: Indian experiences. *Forestry Research and Engineering: International Journal*, 2(6), 306–311. <https://doi.org/10.15406/freij.2018.02.00064>

Sheeren, D., Fauvel, M., Josipovi, V., Lopes, M., & Planque, C. (2016). Tree species classification in temperate forests using Formosat-2 satellite image time series. *Remote Sensing*, 734(8), 1–29. <https://doi.org/10.3390/rs8090734>

Sreenivasulu, G., Jayaraju, N., Pramod Kumar, M., & Prasad, T. L. (2013). An analysis on land use/land cover using remote sensing and GIS – A case study in and around Vempalli, Kadapa District, Andhra Pradesh, India. *International Journal of Scientific and Research Publications*, 3(5), 1-4. Retrieved from [www.ijsrp.org](http://www.ijsrp.org)

Thanh Noi, P., & Kappas, M. (2017). Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using Sentinel-2 imagery. *Sensors*, 18(2), 18. <https://doi.org/10.3390/s18010018>

Torbick, N., Ledoux, L., Salas, W., & Zhao, M. (2016). Regional mapping of plantation extent using multisensor imagery. *Remote Sensing*, 8(3), 236. <https://doi.org/10.3390/rs8030236>

Zhang, X., Friedl, M. A., & Schaaf, C. B. (2009). Sensitivity of vegetation phenology detection to the temporal resolution of satellite data. *International Journal of Remote Sensing*, 30(8), 2061–2074. <https://doi.org/10.1080/01431160802549237>

Zhu, X., & Liu, D. (2015). Improving forest aboveground biomass estimation using seasonal Landsat NDVI time-series. *ISPRS Journal of Photogrammetry and Remote Sensing*, 102, 222–231. <https://doi.org/10.1016/j.isprsjprs.2014.08.014>

## Appendix A: Code for Random Forest Classification in Julia

### Set Image Size and Test Pixel File Location

```
mosaic_ncols = 18447
mosaic_nrows = 18961
k = mosaic_ncols * mosaic_nrows
mosaic_nbands = 30
VNIR_nbands = 4
num_dates = 6
testcol = 9000
testrow = 11000
stack = zeros{Int16, mosaic_ncols, mosaic_nrows, mosaic_nbands};
```

### Initialize Image Stack Array

```
stack = zeros{Int16, mosaic_ncols, mosaic_nrows, mosaic_nbands}
print("\n\nstack array initialized")
print("\nsize of stack: ", size(stack))
```

### Open and Read Image Stack File

```
using ArchGDAL
image_stack_file_name = "input_stack_b.img"
for i = 1:mosaic_nbands
    stack[:, :, i] = ArchGDAL.registerdrivers() do
        ArchGDAL.read(image_stack_file_name) do dataset
            ArchGDAL.read(dataset, i)
        end
    end
end
print("\n", image_stack_file_name, " read\n")
bip_stack = reshape(stack, mosaic_ncols*mosaic_nrows, mosaic_nbands)
println("test pixel (", testcol, ", ", testrow, "):\n\n", bip_stack[mosaic_ncols*(testrow-1)+testcol, :])
using StatsBase
cov_matrix = scattermat(bip_stack)
```

### Build Random Forest Model

```
using DecisionTree
using DelimitedFiles
using Statistics

training_data_file_name = "training_3class_withNDVI_edited_nochar.csv"

num_folds=5
num_subfeatures = -1 #sqrt(num_features)
num_trees = 50
```

```
sampling_proportion = 0.7 #fraction of samples used to train each tree
max_tree_depth = -1 #grown to maximum extent
num_min_leaf_samples = 10
num_min_samples_split = 5
purity_increase_min = 0.0
```

```
godavari = DelimitedFiles.readdlm(training_data_file_name,
    ', ', skipstart=1, Int16)
predictors = godavari[:, 2:31]
target = godavari[:, 1]
```

```
model = build_forest(target,
    predictors,
    num_subfeatures,
    num_trees,
    sampling_proportion,
    max_tree_depth,
    num_min_leaf_samples,
    num_min_samples_split,
    purity_increase_min)
```

```
print("\ncross-validation accuracies:\n\n ")
```

```
accuracy = nfoldCV_forest(target,
    predictors,
    num_folds,
    num_subfeatures,
    num_trees,
    sampling_proportion,
    max_tree_depth,
    num_min_leaf_samples,
    num_min_samples_split,
    purity_increase_min)
```

```
print("for number of trees = ", num_trees, ", mean accuracy = ",
    mean(accuracy), "\n\n")
```

### **Apply Forest and Write File**

```
using Dates
using StatsBase
```

```
print("\n\nApply Forest Start, ", Dates.today(),
    " ", Dates.format(now(), "HH:MM"), "\n\n")
```

```
classification = apply_forest(model, bip_stack)
```

```

# print(classification[mosaic_ncols*(testrow-1)+testcol, 1])
bsq_classification_output = reshape(classification, mosaic_ncols, mosaic_nrows, 1)
# print(bsq_classification_output[testcol,testrow,1])

num_ones,num_twos,num_threes=counts(classification, 1:3)
println("Percent nonforest = ", num_ones / k * 100);
println("Percent plantations = ", num_twos / k * 100);
println("Percent natural forest = ", num_threes / k * 100);

print("\n\noutput array info: ", typeof(bsq_classification_output), " ",
      size(bsq_classification_output), "\n\n")

num_ones,num_twos,num_threes=counts(classification, 1:3)
println("Land cover percentages: ")
println("Percent nonforest = ", num_ones / k * 100)
println("Percent natural forest = ", num_twos / k * 100)
println("Percent plantations = ", num_threes / k * 100)

print("\n\nApply Forest End, ", Dates.today(), " ",
      Dates.format(now(), "HH:MM"), "\n\n")

```

## Write Output File

```
using ArchGDAL; const AG = ArchGDAL
```

```
create_copy_output_file_name = "thematic01_out.img"
classification_output_file_name = "godavari_class_three_sf_b.img"
```

```

AG.registerdrivers() do
  AG.read(create_copy_output_file_name) do dataset
    band1 = ArchGDAL.getband(dataset,1) ## read in 1 band
    ref = AG.getproj(dataset) ## get the project
    new1 = ArchGDAL.read(band1) ## get the data into an array
    geotransform = AG.getgeotransform(dataset) ## get the geotransformation propoerties
    println(geotransform)
    println("projection:")
    println(ref)

    raster = AG.unsafe_create(
      classification_output_file_name,
      AG.getdriver("HFA"),
      width = ArchGDAL.width(dataset),
      height = ArchGDAL.height(dataset),
      nbands = 1,
      dtype = Int16
    )
  end
end

```

```
## assign the projection and transformation parameters
AG.setgeotransform!(raster, geotransform)
AG.setproj!(raster, ref)

## write the raster
AG.write!(
  raster,
  # new1, # image to "burn" into the raster
  bsq_classification_output[:,1], # image to "burn" into the raster
  1, # update band 1
)
AG.destroy(raster)
end
end
```