Regional Mapping of Plantation Extent Using Multisensor Imagery

Nathan Torbick 1,* , Lindsay Ledoux 1, William Salas 1 and Meng Zhao 2

1 Applied Geosolutions, Newmarket, NH 03857, USA; lledoux@appliedgeosolutions.com (L.L.);
wsalas@appliedgeosolutions.com (W.S.)
2 Department of Mathematics and Statistics, University of New Hampshire, Durham, NH 03824, USA;
mbv47@wildcats.unh.edu
* Correspondence: ntorbick@ags.io; Tel.: +1-603-292-1192

Abstract: Industrial forest plantations are expanding rapidly across Monsoon Asia and monitoring extent is critical for understanding environmental and socioeconomic impacts. In this study, new, multisensor imagery were evaluated and integrated to extract the strengths of each sensor for mapping plantation extent at regional scales. Two distinctly different landscapes with multiple plantation types were chosen to consider scalability and transferability. These were Tanintharyi, Myanmar and West Kalimantan, Indonesia. Landsat-8 Operational Land Imager (OLI), Phased Array L-band Synthetic Aperture Radar-2 (PALSAR-2), and Sentinel-1A images were fused within a Classification and Regression Tree (CART) framework using random forest and high-resolution surveys. Multi-criteria evaluations showed both L-and C-band gamma nought $\gamma_0$ backscatter decibel (dB), Landsat reflectance $\rho_\lambda$, and texture indices were useful for distinguishing oil palm and rubber plantations from other land types. The classification approach identified 750,822 ha or 23% of the Taninathryi, Myanmar, and 216,086 ha or 25% of western West Kalimantan as plantation with very high cross validation accuracy. The mapping approach was scalable and transferred well across the different geographies and plantation types. As archives for Sentinel-1, Landsat-8, and PALSAR-2 continue to grow, mapping plantation extent and dynamics at moderate resolution over large regions should be feasible.

Keywords: plantations; rubber; oil palm; PALSAR-2; Sentinel-1; random forest; classification; data fusion; Myanmar; Kalimantan

1. Introduction

The expansion of industrial forest plantations is a critical driver of land cover land use changes in Monsoon Asia. The Food And Agriculture Organization (FAO) estimated 187,086,000 hectares (ha) of forest plantation in 2000 with a rate of 4.5 million new ha/year from 1990 to 2000 [1]. Approximately 79% of the estimated total area was located in Asia. Forest plantations in this report were defined as having a minimum area of 0.5 ha, tree crown cover of at least 10 percent, and a total adult height above 5 m. More recently, a survey conducted by the FAO targeting major producing nations, following up on the Global Forest Resources Assessment, reported plantation extents of 103,728,000 ha and 140,818,000 for 1990 and 2005 worldwide [2]. The report described an augmented approach for data assimilation compared to the 2000 report (e.g., separating productive vs. protective plantations). The report also highlights significant regional and subregional variations between and within years. The discrepancy in these figures emphasize the challenges of producing accurate plantation estimates and the scale of land conversion. Furthermore, the need for tools to accurately characterize the

distribution and dynamics of forest plantations is amplified by international agreements focused on
carbon, conservation, and land management.

Satellite remote sensing is playing an important role in mapping the spatial distribution and
temporal dynamics of forest plantations. A number of studies have used optical satellite images
(Landsat, Spot, and Moderate Resolution Imaging Spectroradiometer (MODIS)) to identify and map
industrial forest plantations [3], specifically rubber [4,5], oil palm [6,7], eucalyptus [8,9], teak [10],
acacia [11,12], and bamboo [13,14]. The main difficulty in mapping industrial forest plantations is the
similar spectral characteristics between natural forests and forest plantations. This is especially evident
when trying to implement traditional classifiers.

One approach to mapping plantations using optical data has been to take advantage of
phenological characteristics unique to a particular species (e.g., rubber, oil palm) in order to separate
them from similar cover types such as natural forest. However, the known limitation of optical data in
the tropics due to cloud cover remains an obstacle for operational (automated) mapping over large
areas and transferability of phenological approaches to different regions or species. Higher temporal
resolution MODIS data has been utilized to circumvent temporally inconsistent moderate scale optical
imagery. A limitation of MODIS imagery is the relatively coarse spatial resolution, as this is an obstacle
in studying fragmented landscapes with patch sizes smaller than the sensor’s spatial resolution.

Some studies have also used synthetic aperture radar (SAR) observations to map rubber and oil
palm plantations [15,16]. The sensitivity of SAR to structural information (biomass, density, vertical
layering) make SAR advantageous especially in the tropics where cloud cover is high. Miettinen and
Liew [17] investigated statistical backscatter signatures of four (wattle, rubber, oil palm, and coconut)
closed canopy plantations in Malaysia and Indonesia. They found statistical differences in backscatter
and HH-HV (horizontal transmitting, horizontal receiving—horizontal transmitting, vertical receiving)
L-band difference able to distinguish oil palm and coconut from other types and attributed this to
the unique structure of palm stands (i.e., branchless stem, flat crown with leaves, open space below
canopy). Highlighted is the need for testing in other conditions, species, and geographic areas.

Multi-sensor and data fusion techniques that integrate multiple types of observations and data
modalities will continue to improve map detail and accuracy. This is partially driven by open
data policies such as the release of Landsat and Phased Array L-band Synthetic Aperture Radar-1
(PALSAR-1) archives and the open distribution of Sentinel-1 and Sentinel-2. By integrating the
strengths of different sensors the limitation of any one sensor can be overcome or complimented,
and additional information can often be derived. Several studies have combined optical and SAR
observations to map rubber plantations [18,19] and oil palm plantations [20]. Recently, Koh et al. [21]
illustrate for a small area how using vegetation indices derived from historical Landsat can help map
timing of events, within a landscape already classified using PALSAR and Landsat, to estimate rubber
plantation stand age.

The overarching goal of this research application was to map plantations at regional scales across
distinct geographies. The specific objectives were to (1) develop multisensor data fusion techniques for
mapping plantations; (2) evaluate transferability across regions and plantation species; and (3) evaluate
new sensors including Sentinel-1 C-band, PALSAR-2 L-band, Landsat-8 Operational Land Imager (OLI)
using 2015 data. The study was carried out in two hot spot regions with rapid plantation development:
West Kalimantan, Indonesia and Tanintharyi, Myanmar.

2. Materials and Methods

2.1. Study Areas

2.1.1. Tanintharyi, Myanmar

The Tanintharyi administrative region is a long narrow body of land adjacent to the Andama Sea
and bordering Thailand (Figure 1). The total land area is 43,344 km² with a population of 1.5 million.
The region is influenced by its tropical monsoon climate and receives upwards of 3000 mm of rain
per year with monthly temperatures above 18 °C. The landscape ranges from coastal villages to high mountain terrain. The FAO estimated a plantation extent increase from 394,000 ha to 849,000 between 1990 and 2005 within Myanmar. In 2013, the volume of timber exports reached 3.3 million m³ with a value of $1.6 billion US, which has tripled in the past decade [22]. As designated historical timber areas are declining, additional natural forested regions are highly susceptible to development. Large-scale land acquisitions for commercial agriculture are a likely outcome as the government begins to evolve and encourage more foreign investment. The Myanmar Forest Department is charged with oversight on timber estates; however, many forested areas remain outside management, policy and coordinated regulation. Tanintharyi, Myanmar has some of the most remote and last remaining “pristine” tracts of tropical forests in the region. Much of the concessions in this ethnically rich and land conflict prone region is driven by oil palm and rubber [22]. Chain of custody and official government data are limited emphasizing the need for robust and transparent Monitoring, Reporting, and Verification (MRV) tools. One report estimated that 28,327 ha of lowland forest was cleared or burned in 2010 and 2011 for oil palm in Tanintharyi concessions, many which are suggested to be located in forest reserves of high conservation value [22,23].

![Figure 1. Tanintharyi, Myanmar and West Kalimantan, Indonesia study areas in tropical South/Southeast Asia with Landsat path rows overlaid and training and validation polygons highlighted in red.](image)

2.1.2. West Kalimantan

West Kalimantan is one of five provinces making up the Indonesian part of Borneo. The boundaries of West Kalimantan nearly follow the rugged terrain around the Kapuas River and border Malaysia. The province has an area of 147,307 km² and nearly 4.5 million people. This study focused on western West Kalimantan plantations hot spots including the entire regencies (kabupaten) of Landak, Sanggau, and Sekadau and portions of Bengkayang, Ketapang, Melwai, Pontianak, Sambas, and Sintang. Kalimantan is characterized by the tropical rainforest climate with average monthly rainfall of 60mm or more. Recent estimates state that half of the world’s oil palm supply comes from Sumatra and Kalimantan and the Agricultural Ministry shows a 600% increase in oil palm between 1990 to 2010. For these reasons, Kalimantan is often considered a major “hot spot” for tropical deforestation and industrial plantations have been identified as the primary driver of extensive loss of peat swamp forest [17,24]. Carlson et al. [7] note that fire is often cited as a driver of deforestation in West Kalimantan; however, they elaborate on that process to report that oil palm was the direct
cause of 27% of total and 40% of peatland deforestation. The FAO estimates that Indonesia as whole contained plantations areas of 2,209,000 ha and 3,399,000 for 1990 and 2005, respectively. Carlson et al. used Landsat and a decision tree approach to map oil palm extent in 1990, calculated to be 90,300 ha to 3,164,000 in 2010 and emphasized the critical role of plantations (extent, distribution, and dynamics) in carbon emissions and driving land use and cover change, which they state is largely undocumented.

2.2. Data Preprocessing

2.2.1. ALOS-2 PALSAR-2

The Advanced Land Observing Satellite (ALOS-2) carries the PALSAR-2 building on the lineage of ALOS-1 PALSAR-1 and Japanese Earth Resources Satellite 1 (JERS-1). ALOS-2 orbits at an altitude of 628 km in a Sun-synchronous pattern with a 14-day revisit cycle. In this study, PALSAR-2 images were collected in Single Look Complex (SLC 1.1) to optimize the complete signal and adjust the effective number of looks considering the ground range resolution, the pixel spacing in azimuth, and incidence angle. Images were co-registered using a cubic convolution cross-correlation approach considering shifts in range and azimuth dependency. A Lee speckle filter was applied to remove spatially random multiplicative noise (speckle). Images were radiometrically calibrated and normalized by eliminating local incident angle effects and antenna gain and spread loss patterns using cosine correction. Terrain geocoding used a Digital Elevation Model (DEM) following the range-Doppler approach to provide gamma nought $\gamma$ dB for the study areas. Local incidence angle $\theta$ and layover/shadow maps were generated for potential post classification processing to adjust for poor data pixels. Twenty-eight (28) and twenty-three (23) single and dual polarization L-band images were used for Tanintharyi and western West Kalimantan, respectively (Appendix 1).

2.2.2. Sentinel-1A

Sentinel-1A carries a C-band imager at 5.405 GHz with an incidence angle between $20^\circ$–$45^\circ$. The platform follows a sun-synchronous, near-polar, circular orbit at a height of 693 km. The 1A platform has a 12-day repeat cycle at the equator. The additional 1B platform planned for launch will increase the repeat coverage by an order of magnitude. Sentinel-1 collects in four modes with different resolutions. The Interferometric Wide (IW) swath collection strategy observes in single and dual polarization VV;VH (vertical transmitting, vertical receiving; vertical transmitting, horizontal receiving) with a 250 km footprint in range direction. All data are freely available from the European Space Agency (ESA) Data Hub. This study utilized SLC and ground range detected (GRD) products that have been focused, multilooked, and projected in ground range. The Sentinel-1 collection strategy began with different observations for SLC and GRD depending on geography, which can be viewed on the ESA Data Hub. Now all regions are operating in the full observation strategy. Images were converted into gamma nought $\gamma$ dB for analyses and mapping. Layover and shadow map were generated for post processing using a DEM. Twenty (20) and fourteen (14) single and dual pol (VV;VH) Sentinel-1 C-band images were obtained for Tanintharyi and West Kalimantan, respectively. Multitemporal imagery for wet and dry seasons were collected to consider the potential influence of phenology in distinguishing rubber and oil palm from other land covers.

2.2.3. Landsat-OLI

Landsat 8 OLI data were collected to provide surface reflectance $\rho_\lambda$ and optical indices to help characterize the landscape. This study used L8SR code to generate surface reflectance and Function of MASK (FMASK) to screen out poor quality optical pixels due to clouds and shadows following lineage of Landsat 5 and 7 preprocessing workflows, i.e., [25–29]. All Landsat data were obtained from United States Geological Survey (USGS) Earth Explorer with phenology and image quality diligently considered during selection. The best available imagery between 2013 and 2015 were selected based on phenology and then mosaicked. A set of well-established indices was also used.
to help classify the landscape. Indices are less sensitive to image-to-image noise, viewing geometry, and atmospheric attenuation making them advantageous over reflectance products in some regard for large area (i.e., multiple scenes over many path rows) and multitemporal mapping. This study used the Normalized Difference Vegetation Index (NDVI; Equation (1)) [30,31], a useful metric of greenness and vigor across a landscape, it is one of the most applied indices for land surface monitoring. The Land Surface Water Index (LSWI; Equation (2)) given its sensitivity to water or equivalent water thickness and leaf moisture has been successfully applied for mapping inundation, forest characteristics, and agricultural landscapes [32,33]. The Normalized Difference Till Index (NDTI; Equation (3)) was used for its sensitivity to residue and crop management practices [34]. The Soil-Adjusted Total Vegetation Index (SATVI; Equation (4)) has demonstrated utility in mapping senescent biomass, ground residue, plant litter, and surface conditions while compensating for varying soil brightness and background artifacts [35].

\[
NDVI = \frac{\rho_{nir} - \rho_{red}}{\rho_{nir} + \rho_{red}}, \quad (1)
\]

\[
LSWI = \frac{\rho_{nir} - \rho_{swir}}{\rho_{nir} + \rho_{swir}}, \quad (2)
\]

\[
NDTI = \frac{\rho_{swir} - \rho_{swir2}}{\rho_{swir} + \rho_{swir2}}, \quad (3)
\]

\[
SATVI = \frac{\rho_{swir} - \rho_{red}}{\rho_{swir} + \rho_{red} + 0.1} \ast (1.1 - \left(\frac{\rho_{swir}}{2}\right)). \quad (4)
\]

2.3. Mapping Approach

Summarizing, our mapping approach centered on generating a suite of complementing inputs and an integrated stack from Landsat-8, PALSAR-2, and Sentinel-1 (Figure 2). The preprocessed imagery and derivatives were stacked into a data cube for analyses and classification. A suite of evaluation techniques was used to investigate classification inputs and accuracies. The fused satellite observations were fed into a classifier to map plantation extent for the study regions.

![Figure 2. Conceptual engineering of multiscale mapping of regional plantation extent.](image_url)
A particular focus was investigating separability of plantations from natural forest, as these tend to be the most easily confused classes based on spectral information. Training and independent validation polygons were developed using multiple high resolution (HR) and field data sources. The first HR data were collected over Kalimantan across 15 sites covering nearly ~65,000 ha in a stratified approach. Very high resolution airborne data covering the visible spectrum (R:G:B) at 50 cm spatial resolution were collected in November (26 November 2014). These data were used to help interpret imagery and generate polygons. Figure 3 illustrates very high resolution airborne data over areas of oil palm, natural forest, and highlights their structural characteristics (e.g., planting spaces, natural forest canopy variation).

The second data source used for interpretation was a limited set of geofield photos. In an effort to promote transparency and improved land cover validation our team has been growing an online archive of field-level photos collected using a GPS-enabled camera (“geofield photos”). All geofield photos are linked to shape files or keyhole markup language (KML) files to store, display, and share photos. KML files use a tag based structure with attributes that allow display. These photos are available for viewing and sharing in Google Earth or any GIS platform at

![Figure 3. Examples from airborne imagery used to help guide development of polygons for training and validation. (A) shows oil palm plantation (top) and natural forest (bottom); (B) shows oil palm plantation (bottom right) and natural forest (top) highlighting natural forest canopy variability; (C) shows oil palm stand with structured planting and cleared ground; (D) shows representative natural forest canopy.](image-url)
www.eomf.ou.edu/photos [36]. At this website, users can search and share a library of global georeferenced field photos for product development and validation. Figure 4 shows a representative geofield photo for creating an agricultural training site. All photos for the study regions were considered for help interpreting imagery (Figure 5) and generating polygons.

Figure 4. Geofield photo (A) of rice agricultural site near Singkawang, West Kalimantan (B).

Figure 5. Co-located examples of (A) airborne imagery; (B) Sentinel-1 (R:VV,G:VH,B:VH²); (C) PALSAR-2 (R:HH,G:HV,B:HV²); and (D) Landsat-8 (R:5, G:4, B:3).
We combined the high resolution imagery and field data with high resolution Google Earth Pro imagery to make final polygons. Google Earth Pro contains time series high resolution data that enables some temporal tracking of landscapes respective of the available dates. A total of 509 polygons containing 1,771,563 30 m pixels were carefully digitized across both study regions (Table 1). A range of plantation ages and landscape conditions (i.e., patch size, slope, interjuxtaposition, distance to urban areas) were included to build a robust calibration and validation data set. Plantation stand age was characterized into three broad classes (young, mixed, mature) based on visual interpretation and examining time-series high resolution imagery in Google Earth Pro.

A suite of derivative indices was generated from both the SAR and optical data. These included common ratios (HH/HV^2) and vegetation indices (NDVI, LSWI, SATVI, and NDTI) that have been found useful in other agroforest mapping studies, i.e., [18,19,32,33,35,37]. Texture indices were also generated in an effort to capture the “uniformity” or homogeneity of plantation canopy, spacing, and structure relative to natural forests. Several studies have found the integration of texture indices useful for mapping forest biometrics, i.e., [37]. Texture indices (Equations (5)–(12)) keyed off gray-level co-occurrence matrix (GLCM) included mean, variance, homogeneity, contrast, dissimilarity, entropy, second moment, and correlation [38]. Next, image statistics for the radar and optical variables were extracted to form a large database for mining, exploration, and model training.

### Table 1. Training data characteristics across both study regions.

<table>
<thead>
<tr>
<th>Class</th>
<th># of Polygons</th>
<th># of Pixels</th>
<th>Min Patch (ha)</th>
<th>Max Patch</th>
<th>Average Patch</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agriculture</td>
<td>87</td>
<td>32,992</td>
<td>0.5</td>
<td>636</td>
<td>33</td>
</tr>
<tr>
<td>Developed</td>
<td>94</td>
<td>37,262</td>
<td>0.4</td>
<td>557</td>
<td>35</td>
</tr>
<tr>
<td>Forest</td>
<td>100</td>
<td>1,103,423</td>
<td>0.8</td>
<td>1211</td>
<td>1102</td>
</tr>
<tr>
<td>Plantation</td>
<td>134</td>
<td>282,215</td>
<td>4.0</td>
<td>3865</td>
<td>192</td>
</tr>
<tr>
<td>Water</td>
<td>94</td>
<td>315,671</td>
<td>1.3</td>
<td>11,370</td>
<td>420</td>
</tr>
</tbody>
</table>

\[
\text{Sum Average (mean)} = \sum_{i,j=1}^{2N_g} \left( p(x+y)(i) \right),
\]

\[
\text{Variance (sum of squares variance)} = \sum_i \sum_j (i - \mu)^2 p(i, j),
\]

\[
\text{Homogeneity (inverse difference moment equation)} = \sum_i \sum_j \left( \frac{1}{1 + (i - j)^2} \right) p(i, j),
\]

\[
\text{Contrast} = \sum_{n=0}^{N_g-1} n^2 \left\{ \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i, j) \right\},
\]

\[
|i - j| = n
\]

\[
\text{Dissimilarity} = \sum_{n=1}^{N_g-1} n \left\{ \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i, j)^2 \right\},
\]

\[
|i - j| = n
\]

\[
\text{Entropy} = -\sum_i \sum_j p(i, j) \log(p(i, j)),
\]

\[
\text{Second Moment} = \sum_i \sum_j (p(i, j))^2,
\]

\[
\text{Correlation} = \frac{\sum_i \sum_j (i)(j)p(i, j) - \mu_x \mu_y}{\sigma_x \sigma_y}\]

p(i,j) is the (i,j)th entry in a normalized gray-tone spatial-dependence matrix; N_g = Number of distinct gray levels in the quantized image; μ_x, μ_y are the means of px and py; σ_x, σ_y are the standard deviations of px and py.

In this study, a classification and regression tree framework was carried out using the multiscale SAR and optical data (Figure 2). The ensemble, machine-learning, random forest algorithm [39]
was employed to classify the remote sensing inputs for mapping plantation extent. A random forest is generated through the creation of a series of Classification and Regression Trees (CARTs) using bootstrapping, or resampling with replacement. Random forest is a flexible and powerful nonparametric technique that many mapping applications have recently implemented for a range of studies including mapping crops [40–42], wetlands [43–45], canopy height [46], algal blooms [47], urban sprawl [48], biomass [49], and many other thematic areas.

For random forests as applied here, a number of decision trees were built and each time a split in a tree is considered, a random sample of \( m (m < p) \) predictors is chosen as split candidates from the full set of \( p \) predictors. In this case \( m: \sqrt{p} \) was applied and this process can be seen as decorrelating the trees to make the average of the resulting trees less variable (reduce the variance). A large tree with more splits may have small bias but lead to higher variance that is challenging to interpret. A small tree with less splits may result in more bias but have lower variance and more straightforward interpretation. With random forest, it is possible to build regression trees that are large, and then prune to determine which subtree gives the lowest tree error rate. Random forest has the ability to treat small and large \( p \) problems, high-order interactions, and correlated predictor variables. An advantage of random forest is the easy integration of multiscale and multimodal input variables, and includes a robust ability to handle large and diverse datasets efficiently.

A complementing set of statistical measures were used to evaluate input variables and determine quality of model outcomes. First, box and whisker plots were generated to simply visualize and assess separability and variability of classes (i.e., plantation, natural forest, urban, water, and agriculture). Second, the Gini index (Equation (13)) was used to help determine the influential remote sensing variables in the CART models, and focused on identifying plantations and characterizing the landscape. Every time a split of a node is made on variable \( m \), the Gini impurity criterion for the two descendent nodes is less than the parent node. Adding up the Gini decreases for each individual variable over all trees in the forest gives a fast variable “importance” that is often very consistent with the permutation importance measure. A small value indicates that a node contains predominantly observations from a single class.

\[
G = \sum_{k=1}^{K} \hat{p}_{mk}(1 - \hat{p}_{mk}),
\]

(13)

where \( \hat{p}_{mk} \) is the proportion of training observations in the \( m \)th region that are from the \( k \)th class. The Gini index is commonly used to measure node purity. A small value indicates that a node contains predominantly observations from a single class; thus, the Gini index can be used to evaluate the importance of a particular split.

Sensitivity (True Positive Rate, TPR) = \[
\frac{True\ Positive}{True\ Positive + False\ Negative},
\]

(14)

False positive rate (FPR) = \[
\frac{False\ Positive}{False\ Positive + True\ Negative},
\]

(15)

Area under the curve (AUC) : \[
AUC = \int_{0}^{1} y(x)dx,
\]

(16)

Precision = True Positive/(True Positive + False Positive),

Recall = True Positive/(True Positive + False Negative).

(17)

(18)

Third, Receiver Operating Characteristic (ROC) curves were generated to help understand classification performance and false positive rates (seen as type I error) against the true positive rate (seen as power) (Equations (14) and (15)). An ROC curve graphs true positive rates \( (y\text{-axis}) \) versus false positive rates \( (x\text{-axis}) \) at a variety of classification thresholds. The true positive rate (tp) is the probability that a predicted value or score, \( S \), for an observation exceeds a classification threshold given
that the observation belongs to the population of interest. The false positive rate (fp) is the probability that the score, $S$, for an observation exceeds the threshold when the observation does not belong to the population of interest. Area under the curve (AUC), precision, and recall are a common approach to further help describe ROC curves (Equations (16)–(18)).

To generalize for polytomous cases, an ROC curve is created for each response level versus the other levels. If there are only two levels, one is the diagonal reflection of the other, representing the different curves based on which is regarded as the “positive” response level. Because the ROC curve summarizes the accuracy of the classification across all possible thresholds, it provides a summary of the overall accuracy of a classification system. A goal is to increase the true positive rate and, at the same time, keep a small false positive rate. Lastly, out of bag (OOB) or “withheld” samples were used to construct error matrices and validate mapping outcomes. Together these metrics provide a robust assessment of mapping capabilities across the different geographies, managements (industrial commercial plantation tracts vs. smallholder plantation plots), and varying landscape conditions (slope, juxtaposition, and heterogeneity).

3. Results and Discussion

3.1. Data Mining

The data mining results indicate that no single sensor provided comprehensive and thorough separation of all the broad classes or the plantations and forest classes. The box and whisker plots showed the PALSAR-2 HV distribution having the most separation between plantation and forested pixels among the three sensors when using SAR $\gamma^\circ$ backscatter and optical $\rho_\lambda$ reflectance observations. PALSAR-2 HV had lower (mean ~4 db lower) backscatter values in both West Kalimantan and Myanmar for plantations compared to natural forest. The L-band wavelength (23.5 cm, 1.25 Ghz) observations penetrate more into the forest canopy and between branches and spaces, compared to the C-band and reflectance measurements, before direct- or double (ground)-bounce scattering or extinction. HH in tropical forests can be dominated by volume and volume-surface scattering while HV can be dominated largely by volumetric scattering [50] due to dense vegetation cover. It is also likely that forest has a higher amount of canopy variability influencing scattering due to significant surface roughness as observed by the cross pol (HV) term. Since structure influences the cross pol term, management (thinning, branch removal, selected harvest, debris removal) is theoretically observable by PALSAR-2 HV if the disturbance is significant enough. Although we note both plantations and “natural” forests can undergo various “managements”.

In young plantations, SAR sensors can receive a relatively higher amount of surface scattering rather than a majority of scattering from trunks and trees or branches and crowns. Therefore, stand age can influence these relationships although variability remains low regardless of age in plantations. As plantations mature, above ground biomass (AGB) increases rapidly. For AGB mapping, the use of L-band [51–53], as well as L- and C-band fusion [54,55], has been shown to be successful although saturation thresholds at given incidence angles are often noted. Typically, the higher the biomass, the stronger the backscatter response at the nominal PALSAR incidence angle, and $150–200$ Mg $\cdot$ ha$^{-1}$ has been noted as HV thresholds where signal to noise begins to wash out [56–61]. We note that the forest class in this study focused on more mature patches (pixels) rather than degraded stands or regrowth patches. Plantations with short life cycles (<5 years) will have a very dynamic range and higher temporal frequency observations will aid in identification.

Sentinel-1 VV and Landsat 8 Shortwave Infrared (SWIR1) had slightly higher backscatter (~1dB) and reflectance values, respectively, in both West Kalimantan and Myanmar for plantations compared to natural forest. This is influenced by crown distribution (density, spacing) and biomass or senescence, respectively, given the particular sensitivity of these bands. It is feasible that newly installed plantations (<2 years old) would have lower volume and much more open ground reducing these separations. Further, the condition of the “background” signal could impact the observations. In some regions,
the influence of peat has been observed in the ability of C-band VV, but not VH, for distinguishing forest types. For example, Quegan and Uryu [62] used Envisat ASAR C-band VV and VH for mapping acacia and found influences of peat in scattering. HH and VH had substantial overlap in box and whisker quantile distributions of plantations and forest as did the reflectance bands of Landsat 8 excluding SWIR2 which had some separation but not as much as SWIR1.

The Landsat texture indices based on NIR had some separation in box and whisker plots between plantations and forest, and more so than SWIR1. NIR variance, contrast, and dissimilarity all had substantial separation with forest values higher than plantations. NIR entropy also had strong separation in box and whisker plots with plantations having higher entropy values. Sentinel-1 VV and VH had similar patterns in box and whisker texture plots with forests having higher values in variance, contrast, and dissimilarity (Figure 6). PALSAR-2 HV and HH also followed the pattern of higher values in variance, contrast, and dissimilarity for forest compared to plantations, although not as pronounced as the other two sensors. Generally the box and whisker plots support the notion that forests have greater variability—in crown size, structure, biomass, density, spacing, stand height, vigor, fractional cover—compared to plantations and the texture metrics help capture this pattern.

![Box and whisker plot showing distributions for plantation and forest for Landsat-8 NIR, PALSAR-2 HH and HV, and Sentinel-1 VV and VH for texture metrics.](image)

**Figure 6.** Box and whisker plot showing distributions for plantation and forest for Landsat-8 NIR, PALSAR-2 HH and HV, and Sentinel-1 VV and VH for texture metrics.

The Gini index showed that VV, VH, SWIR, HV, LSWI, and the texture derivatives homogeneity, dissimilarity, contrast, and entropy were effective for both regions, although not in the same order (Table 2). In Myanmar, the optical variables tended to be recognized as more influential variables while in West Kalimantan the more important variables tended to be dominated by SAR metrics. SWIR was not evident in West Kalimantan, making it less useful in that region as compared to Myanmar. Further, L-band measurements were less important in Myanmar according to the Gini index while proving to be quite important in West Kalimantan. This suggests differences in structure and scale, and potentially other factors such as phenology, varying topography, or background signals, influenced responses.
Table 2. Top Gini index values for (top) West Kalaimantan and (bottom) Tanintharyi using combined Landsat 8, PALSAR-2, and Sentinel-1.

<table>
<thead>
<tr>
<th></th>
<th>HV</th>
<th>VH</th>
<th>VV</th>
<th>VH_Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Greenness</td>
<td>1.759</td>
<td>1.694</td>
<td>1.594</td>
<td>1.499</td>
</tr>
<tr>
<td>VH secondmoment</td>
<td>1.090</td>
<td>1.048</td>
<td>1.030</td>
<td>0.985</td>
</tr>
<tr>
<td>VH entropy</td>
<td>0.974</td>
<td>0.957</td>
<td>0.956</td>
<td>0.832</td>
</tr>
<tr>
<td>NDTI</td>
<td>0.821</td>
<td>0.809</td>
<td>0.795</td>
<td>0.778</td>
</tr>
<tr>
<td>VH correlation</td>
<td>0.747</td>
<td>0.718</td>
<td>0.696</td>
<td>0.690</td>
</tr>
<tr>
<td>VV entropy</td>
<td>1.090</td>
<td>1.048</td>
<td>1.030</td>
<td>0.985</td>
</tr>
<tr>
<td>VV Mean</td>
<td>1.048</td>
<td>1.030</td>
<td>1.000</td>
<td>0.978</td>
</tr>
<tr>
<td>VH Mean</td>
<td>0.985</td>
<td>0.968</td>
<td>0.968</td>
<td>0.937</td>
</tr>
<tr>
<td>VH entropy</td>
<td>0.832</td>
<td>0.821</td>
<td>0.821</td>
<td>0.811</td>
</tr>
<tr>
<td>HH</td>
<td>0.778</td>
<td>0.747</td>
<td>0.747</td>
<td>0.737</td>
</tr>
</tbody>
</table>

The ROC plots (Figures 7 and 8), AUC values (Table 3), and precision and recall graphs (Figure 9) show, as expected, that fusing all three sensors together provides the best overall classification performance. The ROC curves that are closer to the “northwest” corner of the graph represent better classifier performance. The closer the curve comes to the 45-degree diagonal of the ROC space, the less accurate the classifier. The ROC plots show curves using training (T) and withheld (W) outcomes for Landsat-8, PALSAR-2, Sentinel-1, and all three sensors combined. Landsat and Sentinel have similar curves, while forest tends to be more confused when using only Sentinel-1. PALSAR-2 struggles more with agriculture and developed, resulting in lower values in general compared to the other platforms in Myanmar. In West Kalimantan, PALSAR-2 has stronger ROC outcomes and agriculture had the lowest recall and precision value tradeoff. This is likely driven by land use, management, and landscape heterogeneity, as the scale of plantations in West Kalimantan are more industrial (less patchy) as compared to the landscape matrix identified in Tanintharyi.

Table 3. Area under the curve (AUC) values for (top) Tanintharyi and (bottom) West Kalimantan.

<table>
<thead>
<tr>
<th></th>
<th>Landsat 8 OLI</th>
<th>Sentinel-1</th>
<th>PALSAR-2</th>
<th>Fused</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agriculture</td>
<td>Training</td>
<td>Withheld</td>
<td>Training</td>
<td>Withheld</td>
</tr>
<tr>
<td></td>
<td>0.9464</td>
<td>0.9053</td>
<td>0.9864</td>
<td>0.9667</td>
</tr>
<tr>
<td>Developed</td>
<td>0.9919</td>
<td>0.9331</td>
<td>0.9622</td>
<td>0.9039</td>
</tr>
<tr>
<td>Forest</td>
<td>1.0000</td>
<td>0.9733</td>
<td>0.9757</td>
<td>0.8774</td>
</tr>
<tr>
<td>Plantations</td>
<td>0.9835</td>
<td>0.9937</td>
<td>0.9913</td>
<td>0.9713</td>
</tr>
<tr>
<td>Water</td>
<td>1.0000</td>
<td>0.9924</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>Agriculture</td>
<td>0.9966</td>
<td>0.9483</td>
<td>1.0000</td>
<td>0.9822</td>
</tr>
<tr>
<td>Developed</td>
<td>0.9963</td>
<td>0.9874</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>Forest</td>
<td>0.9980</td>
<td>0.9874</td>
<td>1.0000</td>
<td>0.9053</td>
</tr>
<tr>
<td>Plantations</td>
<td>0.9958</td>
<td>0.9378</td>
<td>1.0000</td>
<td>0.9554</td>
</tr>
<tr>
<td>Water</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
</tbody>
</table>
Figure 7. ROC training (T) and withheld (W) curves shown for Taninathryi, Myanmar for classifying agriculture, developed, forest, plantation, and water using (A) Landsat 8 OLI; (B) PALSAR-2; (C) Sentinel-1; and (D) fused/combined imagery.
Figure 8. ROC training (T) and withheld (W) curves shown for West Kalimantan, Indonesia for classifying agriculture, developed, forest, plantation, and water using (A) Landsat 8 OLI; (B) PALSAR-2; (C) Sentinel-1; and (D) fused/combined imagery.
3.2. Mapping

The remote sensing variables ("bands") shown to be effective in distinguishing the cover types according to the data mining results were applied to both regions. In this study, the dominant class was used in the land type map, as determined through the use of random forest and multisensor inputs. In Myanmar, cross validation (withheld random forest model training data) had mean out-of-sample overall and kappa accuracies of 96% and 93%, respectively. The classifier identified 750,822 ha or 23% of the landscape in the plantation class (Figure 10). Comparing these numbers to existing data is challenging given the lack of open and robust record keeping. The extent number is 2–3 times higher than what the Ministry of Agriculture and Irrigation (MOAI) reported for planted oil palm and rubber in 2012 (266,206); however, direct comparisons are not clear given the differences in definitions and lack of access to open and robust data. Figure 8 shows plantation fractional cover and confidence maps for Tanintharyi. The fractional cover map was generated at 10 km cells to visually show percent, and the confidence map (0–100) contains the probability outcomes of random forest that can be used to judge accuracy or possible confusion.

Given the suitable climate for oil palm Tanintharyi has been promoted by the government as a region for industrial plantations. Large oil palm hot pots are evident in southern regions, and rubber across northern districts, with patches existing largely along major transport routes. The majority of plantings tend to spear off “Union Road” in a classic “fishbone” pattern. This is the main transport route in the region, and provides access to districts with dense plantation fractional cover such as Myeik, Dawei, and Kaw Thaung which have many concessions with good access to transport and infrastructure. In 1999, the government had a goal to plant 200,000 ha of oil palm with more than 50 companies receiving concessions [22]. This work suggests plantation extent is larger than current estimates available from FAO or MOAI. With open access to data a more robust and valid cross walk between classifications, concessions and allotments, national parks and reserves, and definitions can be used for transparent accounting and monitoring in Myanmar.
In western West Kalimantan, cross validation had mean out-of-sample overall and kappa accuracies of 98% and 97%, respectively, for the classification model. We emphasize these very high outcomes from the confusion matrix (and thus derived accuracy metrics) are based on withheld cross validation training data from the random forest model. More ground data and transparent comparisons will provide more comprehensive and robust assessments. The classifier identified 216,086 ha or 25% of this landscape study area in the plantation class (Figure 11). Western West Kalimantan has larger patches compared to Tanintharyi, with plantation blocks outside Pontianak and within Pisang, for example, on the order of $10 \times 10$ km of near continuous palm plantation. More inland, many of the larger blocks of plantations tend to be found outside Sintang and the northern half of Sekadau. Topographic influences on the distribution are evident with more extreme topography having intact natural forest cover. Many plantation areas co-occur in peatland regions that are visible when compared to high resolution imagery and existing maps [7]. Again, direct quantitative comparisons to existing maps are challenging given differences in classifications, study area, definitions, and availability of data products.
The mapping outcomes suggest the approach is scalable and transferable across regions and plantations, at least that are dominated by rubber and oil palm. Wall-to-wall maps at national scale at moderate resolution should be feasible. Evaluating across more species, such as acacia, teak, eucalyptus, and bamboo, will provide a more thorough evaluation, as will additional ground truth and additional regions such as Laos and Vietnam. Small plots and roadside plantings are likely under-classified or mixed with forest or agriculture given the spatial resolution of the sensors. Use of Sentinel-1 at finer resolution can provide spatial details on the order of a few trees; however, the fusion with multiple sensors (PALSAR-2, Landsat 8) at moderate resolution (~30 m) provided more robust and accurate maps. Plantations on high relief and complex topography requires effective SAR processing to handle viewing geometry (i.e., incidence angles, layover, foreshortening) and the availability of high resolution DEMs will be required in these cases. With the use of random forest, isolating one or a few critical variables (bands) becomes less critical although still important for explaining what is driving mapping performance and utility. The fusion of L- and C-band with optical data ultimately captures the suite of characteristics that make plantations (oil palm and rubber) unique. Additional temporal information should further improve mapping, although more testing is required. As Sentinel-1B, Sentinel-2A, and PALSAR-2 grow their archives, more opportunity for integration of temporal signals will be available.

4. Conclusions

This study focused on integrating multisensor imagery to map plantation extent across different geographies. The mapping approach was relatively robust across the regions with ability to map oil palm and rubber dominated plantation hot spots with varying landscape characteristics. Landsat 8 OLI, PALSAR-2, and Sentinel-1A each had useful information and the CART approach using random forest transferred well across Tanintharyi and West Kalimantan. PALSAR-2 HV, Sentinel-1 VV and VH, Landsat LSWI and NIR, and texture metrics sensitive to patch contrast and homogeneity were identified as being more useful for distinguishing oil palm and rubber in both regions. The classification approach identified 750,822 ha or 23% of the Tanintharyi, Myanmar, and 216,086 ha or 25% of western West Kalimantan as plantation with very high cross validation accuracy. Potentially, the approach is applicable over larger areas with moderate resolution which should help support monitoring,
reporting, and verification toolsets for understanding plantations and landscapes. The need for open data sharing and robust and transparent plantation assessment tools are evident and will contribute toward understanding land cover land use changes.

Acknowledgments: Support for this work was provided in part by NASA Land Cover Land Use Change Program (NNH12ZDA001N-LCLUC12-2). We thank USGS and ESA for providing Landsat-8 and Sentinel-1 and thank Japanese Aerospace and Exploration Agency (JAXA) Kyoto and Carbon Initiative for access to PALSAR-2. Thanks to Kristin Cichon and Taylor Salas for digitizing polygons and help developing training data. Thanks to Stephen Hagen for access to the airborne data over West Kalimantan as part of a NASA Carbon Management System (CMS) (Grant #: NNX13AP88G) project.

Author Contributions: Nathan Torbick and William Salas conceived and designed the project; Nathan Torbick and Lindsay Ledoux executed image processing and analyses; Nathan Torbick and Meng Zhou performed statistical analysis; Nathan Torbick, William Salas, and Lindsay Ledoux wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest. The funding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

References


47. Song, W.; Dolon, J.M.; Cline, D.; Xiong, G. Learning-based algal bloom event recognition for oceanographic decision support system using remote sensing data. *Remote Sens.* 2015, 7, 13564–13585. [CrossRef]


