Technical Report

Department of Computer Science
and Engineering
University of Minnesota
4-192 Keller Hall
200 Union Street SE
Minneapolis, MN 55455-0159 USA

TR 16-029

Automated Plantation Mapping in Southeast Asia Using Remote Sensing Data

Xiaowei Jia, Ankush Khandelwal, James Gerber, Kimberly Carlson, Paul West, Leah Samberg, Vipin Kumar

August 16, 2016

Added new pdf file
Automated Plantation Mapping in Southeast Asia Using Remote Sensing Data

Xiaowei Jia¹, Ankush Khandelwal¹, James Gerber², Kimberly Carlson³, Paul West², Leah Samberg², and Vipin Kumar¹

¹Department of Computer Science, University of Minnesota, Twin Cities
²Institute on the Environment, University of Minnesota, Twin Cities
³Department of Natural Resources and Environmental Management, University of Hawai’i Manoa

¹jiaxx221@umn.edu, {ankush, kumar}@cs.umn.edu,
²{jsgerber, pcwest, lsamberg}@umn.edu,
³kimberly.carlson@hawaii.edu

Abstract

Plantation mapping is critical for understanding and addressing deforestation, a key driver of climate change and ecosystem degradation. Unfortunately, most plantation maps are limited to small areas for specific years because they rely on visual inspection of imagery. Here we propose an automated approach for annual mapping of plantations using remote sensing data. Due to the heterogeneity of land cover classes, we propose a novel ensemble learning method that simultaneously uses training samples from multiple land cover classes over different years. After the ensemble learning, we further improve the performance by post-processing using a Hidden Markov Model. With the experiments on MODIS data, we demonstrate the superiority of the proposed method over multiple baselines. In addition, we conduct extensive validation by comparing the detected plantation by our approach with the existing datasets developed through visual interpretation by expert observers. Based on the random sampling and the comparison with high-resolution images, the precision (i.e. user’s accuracy) and recall (i.e. producer’s accuracy) of our generated map are around 85.53% and 81.51%, respectively, and the overall accuracy is 95.20%.

Keywords

Remote Sensing; Plantation Mapping; Ensemble Learning
1 Introduction

Global demand for palm oil, driven by demand for edible oils, industrial products, and biofuels (Mukherjee & Sovacool, 2014), is increasing. From 2000 to 2013, global palm oil production grew by 248.06% (according to Food and Agriculture Organization of the United States). Oil palm agriculture is concentrated in Indonesia and Malaysia, which together account for ~85-90% of total global palm oil production. Expansion of oil palm plantations and other types of plantations (e.g. rubber, acacia, etc.) has resulted in substantial tropical deforestation, which releases CO2 from above ground biomass removal and peatland drainage, impacts biodiversity, and affects water quality.

Several companies and governments aim to ensure that the plantations meet rigorous sustainability standards (Scarlat & Dallemand, 2011). For instance, the United Nations Collaborative Programme was launched in 2008 on Reducing Emissions from Deforestation and Forest Degradation in Developing Countries (REDD+). Under a new European Union biofuel policy, any biodiesel import must demonstrate a lifecycle 35% savings in greenhouse gas emissions compared to fossil fuel diesel, and the feedstock cannot be grown in areas with high biodiversity value or high carbon stock. In recent years, many corporations have committed to zero-deforestation supply chains. Companies that produce, trade, and sell palm oil have sought certification including Roundtable on Sustainable Palm Oil (RSPO) (Schouten & Glasbergen, 2011), International Sustainability and Carbon Certification (Moser et al., 2014), and Indonesian Sustainable Palm Oil (ISPO) certification (Paoli et al., 2013). Certification is meant to demonstrate the sustainability of products and supply chains. However, the effectiveness of these diverse policies depends on the monitoring capabilities of different governments and organizations. Hence, scalable and timely monitoring of plantations is essential for successful enforcement of different regulations.
Major recent advances in access to and processing of remote sensing data have enabled semi-automated monitoring of high resolution changes in tree canopy cover over regional to global areas (Hansen et al., 2013; Hansen et al., 2008; Margono et al., 2012). Similarly, recent studies have shown advances in classifying individual crop types, such as soybean and sugarcane in Brazil, using remote sensing data (Rudorff et al., 2010; Rudorff et al., 2011). Yet, discerning natural forests from tree plantations such as rubber, acacia, oil palm, and eucalyptus remains a major challenge. Tree plantations frequently have spectral properties (e.g., greenness) similar to natural forests (Fan et al., 2015; Morel et al., 2011). Certain plantation types have long rotation times between harvests (e.g., oil palm has a ~25 year rotation period), so that clearing and replanting may not be detected in the course of even a long time-series of imagery. As a result, the most widely used global deforestation product (Hansen et al., 2013) defines forests on the basis of structure (tree height and percent canopy cover) and therefore does not differentiate between forest and plantations (Tropek et al., 2014).

For these reasons, many plantation mapping studies have relied on visual interpretation – which takes advantage of human expertise to recognize patterns – in the plantation detection process (Koh et al., 2011; Miettinen, Hooijer et al., 2012; Miettinen, Shi et al., 2012; Petersen et al., 2016). For instance, Miettinen, Shi et al. (2012) manually delineated the industrial plantation based on Landsat 7 images, while Miettinen, Hooijer et al. (2012) first conducted clustering on Moderate Resolution Imaging Spectroradiometer (MODIS) satellite images and then manually checked the properties of each cluster. Similarly, Petersen et al. (2016) first divided the Landsat images into a grid of 20x20 kilometers. Then each gridded structure was visually scanned for plantations with the assistance of forest gain and loss information (Hansen et al., 2013). Koh et al. (2011) clustered ALOS images and visually classified each cluster into basic land cover types (water, forest, plantation, etc.). Then through further manual inspection these clusters were classified into finer-grained land cover types. However, these products
have several limitations for scaling the mapping effort to large regions or globally. First, the visual interpretation may result in both false positives and false negatives due to the quality of images and the observational mistakes. Second, this approach may require multiple people to delineate plantations, likely resulting in inconsistency among observers. Most critically, the human resources needed for large-scale annual digitizing are substantial. Given the challenges of accuracy and inconsistency of observations—as well as the time expense—mapping methods heavily reliant on visual interpretation are not feasible for large regions or completed annually.

A few studies have successfully applied automated machine learning based methods to map plantation agriculture in the tropics. For example, Gutierrez Velez et al. (2013) used a combination of MODIS Enhanced Vegetation Index (EVI) time-series and Landsat to detect forest conversion to oil palm in Peru. Yet, prior research used very simple techniques such as thresholding (Dong et al., 2012; Tabassian et al., 2012; Gutierrez Velez et al., 2013), and the nearest neighbor method (Li & Fox, 2012), which can hardly be adapted to the scenario with multiple land cover types and high-dimensional spectral features. There have been many existing works on multi-class classification (Angulo et al., 2003; Cheong et al., 2004; Wu et al, 2004) and class heterogeneity (Karpatne et al., 2014; Pavlidis et al., 2001). However, these works cannot be directly applied to our problem for several reasons. First, there exists different level of similarity between different land cover types for which we need to aggregate them into high-level classes. Second, the distribution of multiple land covers is extremely skewed, which necessitates an effective sampling process. Finally, the complex high-dimensional feature space in remote sensing data poses a challenge for learning process (Friedman, 1997; Melgani & Bruzzone, 2004). The success of deep learning (Chen et al., 2014; Glorot et al., 2011; Lee et al., 2009) demonstrates its capacity in learning discriminatively from complex feature space and extracting
representative features. Besides, recent works on ensemble learning (Tabassian et al., 2012; Karpatne et al., 2015; Rodriguez-Galiano et al., 2012) distinguish specific land cover type from heterogeneous data with high accuracy. Even after using the most suitable classification machinery, errors in individual annual maps have a compounding effect in context of land cover change detection. Different post processing methods have been proposed in the literature that aim to improve the classification accuracy by using some auxiliary information. For example, Mithal et al. (2013) utilized the Hidden Markov Model to incorporate the transition characteristics of different land cover types and the bias of the classification machinery towards different classes. Utilizing these new machine learning methods offers promise for developing an automated approach for delineating plantations.

Considering the need for accurate and repeatable methods to discriminate tropical tree plantations from natural forests, here we present an automated approach for annually mapping plantations. More specifically, the method uses a combination of remote sensing data, existing visually-delineated plantation maps, and machine learning techniques to distinguish plantations from other land cover types for the Kalimantan region of Indonesia. We compare the accuracy of the results using the new approach with existing products.

2 Dataset and Region of Study

2.1 MODIS data

To map plantations we use the 500 m resolution MODIS data product, which consists of seven reflectance bands (620-2155 nm) collected by MODIS instruments onboard Aqua and Terra satellites. In this product, 8-day composite images are generated from daily images by selecting the per-pixel reflectance value with least noise (i.e. clouds and missing values) from the corresponding 8-day interval.
2.2 Study Region

Our study region is the land area of MODIS tile h29v09 (Fig. 1), which covers much of Kalimantan, Indonesian Borneo, including the province of South Kalimantan, as well as parts of West, Central, and East Kalimantan. Total study area is 328,028 km\(^2\), and contains 1,312,112 locations at 500 m spatial resolution. The whole Kalimantan region has a population of 14,944,742, with yearly average temperature of 21 (low)~32 (high) °C, and monthly average precipitation of 195.2 (Jul)~690.3 (Jan) mm. According to the statistics (Gunarso et al., 2013), Kalimantan region takes account of ~43% oil palm area of Indonesia in 2010.

![Figure 1: Region of Study (MODIS tile h29v09).](image)

2.3 Training Data

Our classification method requires training samples to define the locations and the spectral features of various land cover types. Here, we use two datasets developed by visual interpretation of remote sensing data by human experts to train our classifier. Below we introduce these two datasets.

2.3.1 Tree Plantation Dataset

The Tree Plantation dataset (TP) was developed through visual interpretation of moderate- and high-resolution satellite imagery, and provides the location of tree plantations in selected tropical countries.
Plantations are further categorized as industrial plantation, medium-sized plantation mosaic, small-sized plantation mosaic or very young plantation. Based on the assessment on random stratified sampling (Petersen et al., 2016) conducted as part of the study’s validation process, the dataset’s precision is around 79% (i.e. 79% of the identified tree plantation locations are in fact plantations) and the recall is around 94%. (i.e. 94% of plantations are identified). Hereinafter we will use the term of precision and recall to represent user’s accuracy and producer’s accuracy. In our study region, this dataset provides around 65,100 km² plantation area or 260,483 MODIS pixels in our region of interest.

Table 1: Correspondence between the aggregated classes defined in this study, and high-level classes and land cover types in the RSPO dataset.

<table>
<thead>
<tr>
<th>aggregated</th>
<th>high-level class</th>
<th>land cover type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plantation</td>
<td>Oil palm</td>
<td>Oil Palm Plantation</td>
</tr>
<tr>
<td>Plantation</td>
<td>Timber plantation</td>
<td>Timber Plantation</td>
</tr>
<tr>
<td>Plantation</td>
<td>Agriculture</td>
<td>Rubber Plantation</td>
</tr>
<tr>
<td>Other</td>
<td>Agriculture</td>
<td>Coastal Fish Pond</td>
</tr>
<tr>
<td>Other</td>
<td>Agriculture</td>
<td>Dry Cultivated Land</td>
</tr>
<tr>
<td>Other</td>
<td>Agriculture</td>
<td>Mixed Tree Crops</td>
</tr>
<tr>
<td>Other</td>
<td>Agriculture</td>
<td>Rice Fields</td>
</tr>
<tr>
<td>Other</td>
<td>Built-up</td>
<td>Settlements</td>
</tr>
<tr>
<td>Other</td>
<td>Mining</td>
<td>Mining</td>
</tr>
<tr>
<td>Other</td>
<td>Wasteland</td>
<td>Upland Grassland</td>
</tr>
<tr>
<td>Other</td>
<td>Wasteland</td>
<td>Upland Shrub land</td>
</tr>
<tr>
<td>Other</td>
<td>Wasteland</td>
<td>Swamp Grassland</td>
</tr>
<tr>
<td>Other</td>
<td>Wasteland</td>
<td>Swamp Shrub land</td>
</tr>
<tr>
<td>Other</td>
<td>Water body</td>
<td>Water Bodies</td>
</tr>
<tr>
<td>Other</td>
<td>Disturb forest</td>
<td>Disturbed Mangrove</td>
</tr>
<tr>
<td>Other</td>
<td>Disturb forest</td>
<td>Disturbed Swamp Forest</td>
</tr>
<tr>
<td>Forest</td>
<td>Disturb forest</td>
<td>Disturbed Upland Forest</td>
</tr>
<tr>
<td>Forest</td>
<td>Undisturb forest</td>
<td>Undisturbed Upland Forest</td>
</tr>
<tr>
<td>Forest</td>
<td>Undisturb forest</td>
<td>Undisturbed Swamp Forest</td>
</tr>
</tbody>
</table>

2.3.2 RSPO Dataset

The RSPO dataset, provided by Roundtable on Sustainable Palm Oil (RSPO) (Gunarso et al., 2013), used visual interpretation to delineate industrial oil palm plantations across Indonesia, Malaysia, and
Papua New Guinea in 1990, 2000, 2005, and 2010 (or 2009, depending on the area). In addition, the study digitized all remaining lands into eighteen other land cover types in these eras; the 19 land cover types are aggregated into 9 higher level classes (Table 1, column 2-3). Thus, the dataset provides wall to wall coverage in our region of study. The estimated area and the number of MODIS pixels for each land cover in the study region are depicted in Table 2. This study did not provide an accuracy assessment. Our visual comparison with the high-resolution images from Digital Globe shows that RSPO misses many real plantation areas (low recall). We will show more statistics in Section 5.4 and Section 5.5.

Table 2: Count of pixels by land cover for the years 2000, 2005 and 2009 (column 3-5), and the estimated area \((10^3 km^2)\) of each land cover type for the years 2000, 2005, and 2009 (column 6-8).

<table>
<thead>
<tr>
<th>Full Name</th>
<th>Land Cover</th>
<th>2000</th>
<th>2005</th>
<th>2009</th>
<th>A_{2000}</th>
<th>A_{2005}</th>
<th>A_{2009}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coastal Fish Pond</td>
<td>CFP</td>
<td>5120</td>
<td>5159</td>
<td>6324</td>
<td>1.28</td>
<td>1.29</td>
<td>1.58</td>
</tr>
<tr>
<td>Rubber Plantation</td>
<td>CPL</td>
<td>18398</td>
<td>19813</td>
<td>19741</td>
<td>4.60</td>
<td>4.95</td>
<td>4.94</td>
</tr>
<tr>
<td>Dry Cultivated Land</td>
<td>DCL</td>
<td>44640</td>
<td>57555</td>
<td>86230</td>
<td>11.16</td>
<td>14.39</td>
<td>21.56</td>
</tr>
<tr>
<td>Disturbed Upland Forest</td>
<td>DIF</td>
<td>413561</td>
<td>404786</td>
<td>386326</td>
<td>103.39</td>
<td>101.20</td>
<td>96.58</td>
</tr>
<tr>
<td>Disturbed Mangroove</td>
<td>DIM</td>
<td>6731</td>
<td>6731</td>
<td>6500</td>
<td>1.68</td>
<td>1.68</td>
<td>1.63</td>
</tr>
<tr>
<td>Disturbed Swamp Forest</td>
<td>DSF</td>
<td>81790</td>
<td>83001</td>
<td>66836</td>
<td>20.45</td>
<td>20.75</td>
<td>16.71</td>
</tr>
<tr>
<td>Upland Grassland</td>
<td>GRS</td>
<td>14772</td>
<td>12026</td>
<td>12273</td>
<td>3.69</td>
<td>3.01</td>
<td>3.07</td>
</tr>
<tr>
<td>Mining</td>
<td>MIN</td>
<td>1249</td>
<td>2308</td>
<td>4168</td>
<td>0.31</td>
<td>0.58</td>
<td>1.04</td>
</tr>
<tr>
<td>Mixed Tree Crops</td>
<td>MTC</td>
<td>6944</td>
<td>7657</td>
<td>7995</td>
<td>1.74</td>
<td>1.91</td>
<td>2.00</td>
</tr>
<tr>
<td>Oil Palm Plantation</td>
<td>OPL</td>
<td>27948</td>
<td>42572</td>
<td>101806</td>
<td>6.99</td>
<td>10.64</td>
<td>25.45</td>
</tr>
<tr>
<td>Rice Fields</td>
<td>RCF</td>
<td>28697</td>
<td>29416</td>
<td>30419</td>
<td>7.17</td>
<td>7.35</td>
<td>7.60</td>
</tr>
<tr>
<td>Upland Shrub land</td>
<td>SCH</td>
<td>288002</td>
<td>294930</td>
<td>258922</td>
<td>72.00</td>
<td>73.73</td>
<td>64.73</td>
</tr>
<tr>
<td>Settlements</td>
<td>SET</td>
<td>2776</td>
<td>2839</td>
<td>2840</td>
<td>0.69</td>
<td>0.71</td>
<td>0.71</td>
</tr>
<tr>
<td>Swamp Grassland</td>
<td>SGR</td>
<td>16713</td>
<td>13887</td>
<td>13525</td>
<td>4.18</td>
<td>3.47</td>
<td>33.8</td>
</tr>
<tr>
<td>Swamp Shrub land</td>
<td>SSH</td>
<td>98669</td>
<td>103509</td>
<td>108240</td>
<td>24.67</td>
<td>25.88</td>
<td>27.06</td>
</tr>
<tr>
<td>Timber Plantation</td>
<td>TPL</td>
<td>12008</td>
<td>12531</td>
<td>12117</td>
<td>3.00</td>
<td>3.13</td>
<td>3.03</td>
</tr>
<tr>
<td>Undisturbed Upland Forest</td>
<td>UDF</td>
<td>136217</td>
<td>115656</td>
<td>97007</td>
<td>34.05</td>
<td>28.91</td>
<td>24.25</td>
</tr>
<tr>
<td>Undisturbed Swamp Forest</td>
<td>USF</td>
<td>88069</td>
<td>77928</td>
<td>71035</td>
<td>22.02</td>
<td>19.48</td>
<td>17.76</td>
</tr>
<tr>
<td>Water Bodies</td>
<td>WAB</td>
<td>19808</td>
<td>19808</td>
<td>19808</td>
<td>4.95</td>
<td>4.95</td>
<td>4.95</td>
</tr>
</tbody>
</table>

3 Method

Plantation mapping requires differentiation between plantations and multiple non-plantation land cover types, which may range from bare soils to primary intact forests. Here we propose to aggregate the land cover types and learn a three-class ensemble. We will also explore the sampling and learning model.
3.1 Ensemble Learning Method

The RSPO dataset provides excellent training data for our ensemble learning method, because of its large number of classes, multiple time points, wall-to-wall coverage, and high precision. Given the land cover types defined by RSPO, we need to aggregate them into several classes to train a learning model. Although the detailed land cover taxonomy can obtain the maximum information about land cover change, the existing maps do not achieve high accuracy on fine-grained land covers. On the contrary, if we aggregate all the land cover types other than plantation into “non-plantation” class, the high heterogeneity within the “non-plantation” class will hamper the learning performance and increase the risk of misclassification (Karpatne et al., 2015). While distinguishing between plantation and forest is challenging due to similar per-pixel spectral properties, other land cover types such as urban areas and annual crops are typically spectral dissimilar from forests and plantations, and are therefore easier to distinguish from plantation. Besides, we are interested in our method’s ability to distinguish forests from plantations, and therefore we retain the “forest” class separate from other land use types. Based on these considerations, we aggregate all land cover types into three classes: “forest”, “plantation” and “other”, as described in the first column of Table 1. The “forest” class contains both undisturbed forest as well as logged or degraded yet natural forests.

To discover the discriminative knowledge between each pair of classes, we learn a multi-classification ensemble. Specifically, we train three binary classifiers using a one vs one strategy: plantation versus forest (P-F), forest versus other (F-O) and other versus plantation (O-P). In this way each classifier focuses on exploiting the discriminative knowledge between a specific pair of classes while ignoring the other class. Hence this separate learning strategy can assist in reducing the heterogeneity and improving the learning performance.
3.1.1 Aggregating Result in Prediction

After training the three classifiers separately, we then discuss how to aggregate the predicted results from them to make final decision. Since each binary classifier focuses on differentiating between a specific pair of classes, eight possible combinations represent potential outcomes from the three classifiers. Based on the separate predictions from each classifier, we assign the aggregated prediction result as the majority class label. For instance, if both P-F and O-P classifiers predict a test location as “plantation”, then we will label this test location as “plantation” regardless of the prediction of F-O classifier. When the three binary classifiers generate mutually different labels the test sample will be assigned to the “Unknown” (U) class. We summarize the relationship between each individual prediction and the aggregated prediction in Table 3.

According to the first two rows in Table 3, we guarantee that the detected locations are more likely to be plantation when separately compared to either forest or other land cover types. Besides, we will not select the locations that are marked as plantation by only one of P-F or O-P. In this way, we can improve the precision of the prediction result.

Table 3 Aggregation of prediction from pair-wise classifiers.

<table>
<thead>
<tr>
<th>P-F</th>
<th>F-O</th>
<th>O-P</th>
<th>Aggregation</th>
</tr>
</thead>
<tbody>
<tr>
<td>P</td>
<td>F</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>P</td>
<td>O</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>O</td>
<td>F</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>P</td>
<td>F</td>
</tr>
<tr>
<td>P</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>F</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>P</td>
<td>F</td>
<td>O</td>
<td>U</td>
</tr>
<tr>
<td>F</td>
<td>O</td>
<td>P</td>
<td>U</td>
</tr>
</tbody>
</table>
3.2 Sampling Strategy

Aggregated land cover classes contain multiple land cover types, and relative area under each land cover type varies widely. For instance, in the RSPO dataset, settlement area is much smaller than grassland area. If we uniformly sampled from each aggregated non-plantation land cover class for supervised classification, the resulting map would likely be dominated by the land cover types with large relative area, such as forest and wasteland (Sun et al., 2009). As a result, minority classes such as settlement might be misclassified as plantation and this would reduce the precision of our classification. To address this issue, we simultaneously sample and learn from multiple land cover classes. Specifically, we uniformly take plantation samples from the defined “plantation” class. Then for the “forest” class and “other” class, we randomly sample equal amount of pixels from each sub-class within the aggregated class.

Due to temporal heterogeneity in MODIS data, a model developed for a specific year can result in a poor prediction when applied on another year. To overcome this limitation, we sample simultaneously from multiple years. Given available RSPO data at multiple time steps \( \{t_1, t_2, \ldots, t_m\} \), if a location is in a given land cover class at both \( t_i \) and \( t_{i+1} \), we assume it remains in this class for each year between \( t_i \) and \( t_{i+1} \). In this way we can take MODIS data samples from multiple years for the “plantation”, “forest” and “other” class. The collection of samples from multiple years enables the training of a global model which can be applied over an entire observation period.

Specifically, according to the availability of RSPO dataset, we first divide the 2001 to 2014 period into three intervals. For the first two intervals, i.e. 2001-2005 and 2006-2009, we are only confident with the locations that are labeled as same land cover types at both end years. For the 2010-2014 interval, we are confident with the locations as land cover type \( k \) if they are labeled as land cover type \( k \) in 2009 and
showing little change of Enhanced Vegetation Index (EVI) during 2009-2014. Then we will utilize TP dataset to further prune the confident samples. In particular, for the “plantation” class, we keep only those confident samples that are also included in the TP dataset. As mentioned earlier, TP dataset has high recall, i.e., it includes a vast majority of true plantations. If RSPO dataset labeled a location to be plantation that is not part of TP dataset, then this location would probably be a potential false positive in RSPO dataset. For the “forest” class and “other” class, we use the same characteristic of TP dataset to prune the confident samples. Specifically, we keep only those confident samples that are not included in the TP dataset.

3.3 Learning model

After gathering training samples, we train a Deep Belief Network (DBN) (Bengio et al., 2007) to learn each of the three binary classifiers. DBN is effective in exploiting the latent relationship among input features and extracting high-level representative features. As mentioned in Section 4.2, the samples of each class can be associated with multiple land cover types, which requires that the model effectively transforms the spectral features into a subspace where the land cover types within each class are close. To achieve this goal, the model should be able to automatically extract the features that best discriminate between classes. We take advantage of the spectral and temporal richness of our data by exploiting reflectance values from multiple bandwidths, as well as information collected from different periods of a year, to determine land cover. Given its efficacy in extracting latent representative features
via the unsupervised pre-training phase and the supervised fine-tuning process, a DBN is an excellent tool for such land cover classification.

The main component of DBN is Restricted Boltzmann Machine (RBM), an undirected graphical model structured as a fully connected bipartite graph between two layers of binary variables, visible variables $V \in R^N$ and hidden or latent variables $V \in R^M$, as shown in Fig. 2. In our problem $V$ represents concatenation of reflectance values over multiple composite images in a specific year, which has the length of $N = N_b \times N_d$. $N_b$ and $N_d$ denote the number of band and the number of selected days (or composite images) in a year. In RBM, the joint distribution of $V$ and $H$ is defined by an energy based distribution:

$$P(V, H) = \frac{1}{Z} \exp(-E(V, H)), \quad (1)$$

where $Z$ represents the normalizer, and the energy function $E(V, H)$ is defined as:

$$E(V, H) = -\sum_{ij} V_i W_{ij} H_j - \sum_j B_v V_j - \sum_i B_h H_i, \quad (2)$$

where $W \in R^{M \times N}$ stands for the weight matrix that connects to the visible variables and the hidden variables, and $B_v \in R^N$ and $B_h \in R^M$ represent the biases for the $V$ and $H$, respectively.

Since there are no connections between the variables in the same layer, the conditional probability becomes fully factorial:

$$P(H_i = 1|V) = \sigma(W_i V + B_h),$$

$$P(V_i = 1|H) = \sigma(W^T_j H + B_v), \quad (3)$$

where $W_i$ and $W_j$ represent the $i^{th}$ row and $j^{th}$ column of $W$, respectively. $\sigma(\cdot)$ denotes the sigmoid function.
The objective of RBM is to maximize the likelihood of visible variables. Since the term $Z$ in Eq. 1 involves the summation over exponential terms, it is computationally intractable to directly estimate the derivative. Instead, we perform alternating Gibbs sampling and use the approximated derivative to update the model parameters (Hinton, 2002).

While RBM offers opportunity to extract latent features, one single RBM may fail to reveal the truly discriminative features due to the limited ways of combination of input features. Therefore, researchers usually stack multiple RBM layers to form DBN. DBN enables the learning of more representative features since the ways of combination increase exponentially with the number of layers. The DBN model can be trained in a greedy (layer-wise) fashion (Bengio et al., 2007).

3.4 Filtering

During the analysis of our ensemble classification machinery, we found a special set of locations can be easily mixed with plantation since they have similar characteristics, e.g. they both have crossing roads among the trees, and most of these samples belong to the wasteland class defined by RSPO. Since these locations are similar with plantations they are likely to be misclassified as plantations, but with less confidence than true plantations. One reason that leads to the classification error is the lack of such training samples when conducting three-class classification. Even though we collect equal number of samples from each land cover type to obtain a rich training set, samples within a given land cover type are selected randomly. Since these special locations are present in very small fraction in the wasteland class, we can only get limited samples on this type. When these limited samples are mixed with other land cover types in the “other” class, the training process can be dominated by the other samples that
are more distinguishable from plantation. Besides, these special locations may be treated as outliers since they are more similar to the “plantation” class than the “other” class.

To solve this problem, we train a separate classifier between wasteland and plantation (W-P). Then we utilize this classifier to filter the detected plantation locations obtained from the three-class ensemble classification step. Specifically, we apply the W-P classifier only on locations that are classified as plantation by the 3-class ensemble classification model. Previously detected plantation locations that get labeled as wasteland by the W-P classifier are filtered out. In this way we are able to remove false positive locations that are detected with less confidence and are similar with wasteland, thereby improving the precision of the resulting classification.

3.5 Post-processing

Due to the natural disturbance and the temporal variation in remote sensing data, conducting separate annual predictions results in classification errors and inconsistencies among years. Here, we use a Hidden Markov Model based post processing scheme to improve the quality of the classification labels. Different land cover types have different transition or conversion characteristics. For instance, in most plantation-intensive areas, forest is converted to plantation land cover, while plantations are highly unlikely to be converted back to forest during our study period.

Consider a yearly sequence of \{forest, forest, plantation, plantation, forest, plantation\}. Here, the third “forest” is highly likely to be a classification error and should be fixed to “plantation”. To capture this transition relationship, we utilize Hidden Markov Model (HMM), which models the transition probability among latent states by the transition matrix \(T\) and the mapping relationship between the latent state and the observed class by the emission matrix \(E\). In particular, \(T_{ij}\) represents the transition
probability from state $i$ and state $j$, and $E_{ik}$ denotes the emission probability from state $i$ to the observed class $k$.

In our problem each latent state represents a real land cover type and each observed class $k \in \{"plantation", "forest" or "other"\}$. We initialize the transition matrix using the visually delineated RSPO land cover dataset which is available at multiple time steps $\{t_1, t_2, ..., t_m\}$. Specifically, $T_{ij}$ is initialized as the proportion of locations in land cover $i$ at any time step from $\{t_1, t_2, ..., t_{m-1}\}$ to be converted to land cover $j$ at next time step. On the other hand, each entry in emission matrix $E_{ik}$ represents the probability for a real land cover type $i$ to be classified as class $k \in \{"plantation", "forest" or "other"\}$. In this way the emission matrix $E$ can capture the confusion between land cover classes. With the obtained transition matrix and emission matrix, we can fix the yearly prediction on each location via Viterbi algorithm (Forney, 1973).

Besides HMM, we also conduct post-processing from spatial perspective. It is known that plantation lands are frequently developed by cutting or burning trees near existing plantations. Therefore, we implement the spatial filtering process in the following steps. First, we conduct spatial clustering by finding all the connected components of plantation locations. Then for each cluster, if the cluster size is less than a threshold $\delta$, we will consider this cluster as false positive and remove this cluster. In our implementation we set $\delta = 30$, which is equivalent to allow the minimum plantation area of $7.5 \ km^2$.

4. Validation

Although the TP and RSPO datasets are not perfect datasets as they contain errors of commission and omission, we treated them as “truth” for purposes of model building and classification. But for
validation, we use these datasets for comparative analysis while being aware of their error characteristics so that a more accurate evaluation can be performed. RSPO has relatively low errors of commission but it has high errors of omission. In other words, locations detected as plantations by RSPO can be trusted. An algorithm’s performance can be evaluated by quantifying how many of these trusted locations are detected by the given algorithm. Specifically, RSPO dataset can be used to evaluate the recall of an algorithm with respect to the locations detected by RSPO. Note that the recall obtained using RSPO dataset will not be exactly same as the true recall but this is the best estimate of recall that can be achieved using these datasets. Since our proposed method generates yearly plantation maps, we can validate annual performance. Besides, we will measure the overall performance of plantation detection from 2001 to 2014. Specifically, we measure the performance in terms of recall for each year from 2001 to 2009, based on the confident locations of forest, other, and plantation described in Section 3.2. Since these confident locations are obtained based on the available years of RSPO and cannot cover all the plantation locations for each intermediate year, we cannot well estimate the precision on these years. Instead, since TP dataset has a high recall, i.e. very low error of omission, the detected locations outside TP are highly likely to be false positives and hence we measure the overall precision using the TP dataset. Similarly, we measure the overall recall using the RSPO dataset on 2009. The overall performance is measured based on all the detected plantation locations through 2001 to 2014.

4.1 Comparison of Learning Strategies and Post-Processing Techniques

To evaluate the effectiveness of our proposed ensemble learning model and the sampling strategy, we compare the following classification methods to the core method described above (termed PALM for Plantation Analysis by Learning from Multiple Classes):
**Bin:** In this method, we train a binary classifier between plantation and non-plantation, where non-plantation merges forest and other classes. Wasteland filtering is applied to improve precision.

**NonF:** In this strategy, we conduct three-class classification but without wasteland filtering.

**UniS:** Here, we uniformly sample from the entire “Other” class rather than take equal amount of samples from each land cover type.

**SVM:** Instead of DBN, we implement our ensemble learning strategy using Support Vector Machine with RBF kernel.

In the post-processing steps, we utilize the 19 land cover types defined in RSPO for HMM. To analyze the efficacy of post-processing, here we compare the *PALM* method to the following variants:

**NonP:** In this baseline we implement the proposed learning method without using post-processing process.

**H9:** Here we conduct HMM post-processing based on 9 high-level classes provided in RSPO dataset.

### 4.2 Comparison with TP and RSPO using Digital Globe

As mentioned earlier, both TP and RSPO datasets have errors. Here we would like to further evaluate the quality of our detection by investigating the errors made by the proposed approach with respect to these datasets. We focus on the RSPO map on 2009, which is almost a subset of TP (92% of RSPO map is included in TP). Specifically, we will investigate the difference between our detected plantation locations (through 2001 to 2014) and the two available datasets. We consider the difference in the following four regions separately:

- **R1** - plantations detected by *PALM* and TP, but missed by RSPO.
- **R2** - plantations detected by TP, but missed by *PALM* and RSPO.
- **R3** - plantations detected by *PALM*, but missed by TP.
R4 – plantations detected by RSPO, but missed by PALM.

We will first show some cases for each of these regions. Based on these cases we wish to demonstrate: 1) we can detect more real plantations than RSPO (through the cases in R1), 2) we can avoid including the false positives from TP (through the cases in R2), 3) we can detect some real plantations outsides TP (through the cases in R3), and 4) RSPO contains false positives and we can avoid including them (through the cases in R4).

After the case study we will quantify our analysis by comparing the random samples from each region to the high-resolution images from Digital Globe. In particular, we wish to obtain the fraction of samples in \{R1, R2, R3, R4\} that are real plantations. Based on the quantified results on \{R1, R2, R3, R4\}, we can estimate the confusion matrix between our detected plantations and the reference of Digital Globe.

5. Results and Discussion

5.1 Plantation map and basic statistics

We first show our plantation map of the study region on 2014 in Fig. 3. Here the detected plantation locations are marked in brown color.

Figure 3: The plantation map on 2014. The plantation locations are marked in brown color.
Then we show the annual plantation area on our study region from 2001 to 2014 in Fig. 4. Based on the statistics of plantation area, the annual growth from 2001 to 2014 is around 9.57%.

![Figure 4: The annual plantation area on the study region from 2001 to 2014.](image)

5.2 Comparison with different learning strategies

Table 4: Comparison to different learning strategies - yearly recall from 2001 to 2009, overall precision and overall recall.

<table>
<thead>
<tr>
<th>Method</th>
<th>2001</th>
<th>2002</th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
<th>2008</th>
<th>2009</th>
<th>Overall precision</th>
<th>Overall recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bin</td>
<td>0.733</td>
<td>0.782</td>
<td>0.804</td>
<td>0.813</td>
<td>0.821</td>
<td>0.766</td>
<td>0.781</td>
<td>0.798</td>
<td>0.823</td>
<td>0.835</td>
<td>0.843</td>
</tr>
<tr>
<td>NonF</td>
<td>0.780</td>
<td>0.823</td>
<td>0.841</td>
<td>0.853</td>
<td>0.866</td>
<td>0.804</td>
<td>0.819</td>
<td>0.834</td>
<td>0.856</td>
<td>0.798</td>
<td>0.859</td>
</tr>
<tr>
<td>UniS</td>
<td>0.752</td>
<td>0.806</td>
<td>0.828</td>
<td>0.849</td>
<td>0.799</td>
<td>0.814</td>
<td>0.832</td>
<td>0.847</td>
<td>0.810</td>
<td>0.783</td>
<td>0.862</td>
</tr>
<tr>
<td>SVM</td>
<td>0.544</td>
<td>0.854</td>
<td>0.720</td>
<td>0.760</td>
<td>0.780</td>
<td>0.736</td>
<td>0.749</td>
<td>0.756</td>
<td>0.760</td>
<td>0.736</td>
<td>0.643</td>
</tr>
<tr>
<td>PALM</td>
<td>0.759</td>
<td>0.816</td>
<td>0.842</td>
<td>0.853</td>
<td>0.867</td>
<td>0.810</td>
<td>0.823</td>
<td>0.837</td>
<td>0.858</td>
<td>0.846</td>
<td>0.868</td>
</tr>
</tbody>
</table>

Here we show the performance of PALM and other learning strategies in Table 4. It is noteworthy that the post-processing step has been applied on all the learning strategies in order to compare the best performance of each strategy. Besides, yearly recall values have been calculated on the confident locations on each year and overall recall is measured on RSPO dataset on 2009, as explained in Section 4. We observe that the binary classification Bin performs not as good as PALM due to the strong heterogeneity within the non-plantation class. Moreover, the comparison between NonF and PALM shows the effectiveness of wasteland filtering in improving precision. On the other hand, the
performance of UniS is unsatisfactory since the training is dominated by the land cover types with large population, and ignores the small classes that are similar to plantation. In this way the trained classifier is highly likely to misclassify these small classes as plantation, and consequently leads to low precision. Furthermore, we can observe that PALM outperforms SVM by a considerable margin due to the effectiveness of DBN in learning from complex feature space.

5.3 Comparison of post-processing steps

We show the performance of PALM and the baselines in Table 5. The comparison between PALM and NonP demonstrates the effectiveness of post-processing. Besides, the HMM using 19 land cover classes outperforms the learning on 9 classes. This is mainly because the 19 classes can better define the latent state space in HMM and more accurately model the transition process.

Table 5: Comparison to different post-processing strategies - yearly recall from 2001 to 2009, overall precision and overall recall.

<table>
<thead>
<tr>
<th>Method</th>
<th>2001</th>
<th>2002</th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
<th>2008</th>
<th>2009</th>
<th>Overall precision</th>
<th>Overall recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>NonP</td>
<td>0.767</td>
<td>0.757</td>
<td>0.786</td>
<td>0.771</td>
<td>0.828</td>
<td>0.740</td>
<td>0.743</td>
<td>0.761</td>
<td>0.815</td>
<td>0.750</td>
<td>0.833</td>
</tr>
<tr>
<td>H9</td>
<td>0.721</td>
<td>0.789</td>
<td>0.818</td>
<td>0.839</td>
<td>0.860</td>
<td>0.806</td>
<td>0.822</td>
<td>0.841</td>
<td>0.866</td>
<td>0.825</td>
<td>0.880</td>
</tr>
<tr>
<td>PALM</td>
<td>0.759</td>
<td>0.816</td>
<td>0.842</td>
<td>0.853</td>
<td>0.867</td>
<td>0.810</td>
<td>0.823</td>
<td>0.837</td>
<td>0.858</td>
<td>0.846</td>
<td>0.868</td>
</tr>
</tbody>
</table>

5.4 Validation against TP and RSPO

Using the land use classification from PALM, we compute the confusion matrix with respect to both TP and RSPO, as shown in Table 6. Here we utilize the generated plantation map in 2014 when comparing with TP while for RSPO we use the generated plantation map in 2009. From Table 6 we can observe that our generated plantation map is almost included in TP dataset, but misses some locations of TP dataset. On the other hand, the generated map covers most plantation areas of RSPO dataset, but also detect more locations outside RSPO dataset.
Table 6: The confusion matrix in percentage with respect to Tree Plantation and RSPO dataset.

<table>
<thead>
<tr>
<th>PALM</th>
<th>Plantation</th>
<th>Non-plantation</th>
<th>Total</th>
<th>TP</th>
<th>Plantation</th>
<th>Non-plantation</th>
<th>Total</th>
<th>RSPO</th>
<th>Plantation</th>
<th>Non-plantation</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>PALM</td>
<td>Plantation</td>
<td></td>
<td></td>
<td></td>
<td>Plantation</td>
<td></td>
<td></td>
<td></td>
<td>Plantation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PALM</td>
<td>Non-plantation</td>
<td></td>
<td></td>
<td></td>
<td>Non-plantation</td>
<td></td>
<td></td>
<td></td>
<td>Non-plantation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>11.99</td>
<td>2.18</td>
<td>14.17</td>
<td>6.46</td>
<td>5.32</td>
<td>11.78</td>
<td></td>
<td></td>
<td>18.54</td>
<td>81.46</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>6.55</td>
<td>79.29</td>
<td>85.83</td>
<td>1.30</td>
<td>86.92</td>
<td>88.22</td>
<td></td>
<td></td>
<td>7.76</td>
<td>92.24</td>
<td>100</td>
</tr>
</tbody>
</table>

5.5 Case Studies

Here we wish to understand whether our generated plantation map can achieve a better balance between precision and recall than TP and RSPO. Specifically, we will focus on the four regions {R1, R2, R3, R4} defined in Section 4.2. For simplicity here we consider the plantation locations detected by our approach as those identified from 2001 to 2014. We first conduct several case studies and then show the quantified result by analyzing a set of sampled locations in each region.

Cases in R1 and R2: To analyze the cases in R1 and R2, we show three cases in Fig. 5. Here the red color represents the locations in R1, and the blue color represents the locations in R2. We show the high-resolution images (via Digital Globe) corresponding to Fig. 5 (a), (c) and (e) in Fig. 5 (b), (d) and (f), respectively. According to the high-resolution images, the red colored region in Fig. 5 (a) is a real plantation area, but is missing from RSPO dataset. Since RSPO dataset provides map on 2009, it is also possible that these missing locations grow into plantations after 2009. As for the blue colored region in Fig. 5 (c), which is included by Tree Plantation dataset but not detected by our method, we can clearly see from the high-resolution image that it is not real plantation. In Fig. 5 (e) we show an area with locations in R1 and R2. From the high-resolution image in Fig. 5 (f), we observe that the proposed method can well detect the boundary between real plantation and non-plantation area.
With these examples in R1 and R2, we demonstrate that our proposed method can detect the real plantation locations that are missing from RSPO dataset while also avoiding the locations that are mistakenly detected by Tree Plantation dataset.

Figure 5: The examples of R1 and R2. (a) The selected locations on Google Earth terrain image. (b) The corresponding high-resolution image.

Cases in R3: Now we show several examples for the locations that are detected by our approach but missed by Tree Plantation dataset in yellow color, as displayed in Fig. 6 (a) and (c). By using the high-
resolution images, we can clearly see that they are real plantation. In this way we demonstrate that our method can detect real plantations outside TP.

Figure 6: The examples of R3. (a) (c) The selected locations on Google Earth terrain images. (b) (d) The corresponding high-resolution images.

Figure 7: The examples of R4. (a) (c) The selected locations on Google Earth terrain images. (b) (d) The corresponding high-resolution images.
**Cases in R4:** In this part we show several examples of locations that are detected by RSPO dataset but missed by our approach. We show these locations (R4) in green color. Locations in R4 are usually adjacent to or along the boundary of the locations that are included by both RSPO dataset and our approach, which are marked by magenta color. We show two large patches of R4 locations in Fig. 7 (a) and (c). According to the corresponding high-resolution images, these locations are not real plantation. Therefore, we demonstrate that RSPO wrongly identifies some areas as plantation while our proposed approach can avoid making these mistakes.

To better show the effectiveness of our method and to quantify the performance, we randomly sample locations from each region \{R1, R2, R3, R4\}, and manually validate the correctness of these locations by comparing to high-resolution images. The result is shown in Table 7. We can observe that most detected locations in R1 are real plantations and a small portion of locations in R2 are real locations. Moreover, although RSPO has high precision and TP has high recall, we can observe that both R3 and R4 have around 40% plantation locations.

**Table 7: Analysis of random samples by comparing to TP and RSPO.**

<table>
<thead>
<tr>
<th>Region</th>
<th>samples</th>
<th>plantation</th>
<th>percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>375</td>
<td>307</td>
<td>81.87%</td>
</tr>
<tr>
<td>R2</td>
<td>200</td>
<td>78</td>
<td>39.00%</td>
</tr>
<tr>
<td>R3</td>
<td>364</td>
<td>156</td>
<td>42.86%</td>
</tr>
<tr>
<td>R4</td>
<td>358</td>
<td>150</td>
<td>41.90%</td>
</tr>
</tbody>
</table>

According to Table 7 and the number of samples in each defined region \{R1, R2, R3, R4\}, we can estimate the confusion matrix with respect to the high-resolution images from Digital Globe. Specifically, we assume that the locations that lie in the intersection of our detection, TP, and RSPO are correct, and the fraction of plantations in \{R1, R2, R3, R4\} is indicated in Table 7. Therefore, we can estimate the total number of real plantation locations and the number of real plantation locations.
detected by our approach. We show these values in confusion matrix in Table 8. According to the table, the precision and recall of our generated map is 85.53% and 81.51%, respectively, and the overall accuracy is 95.20%.

Table 8: The estimated confusion matrix in percentage between the generated map in 2014 and the digital globe high-resolution image.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Plantation</th>
<th>Non-plantation</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>PALM Plantation</td>
<td>12.12</td>
<td>2.05</td>
<td>14.17</td>
</tr>
<tr>
<td>Non-Plantation</td>
<td>2.75</td>
<td>83.08</td>
<td>85.83</td>
</tr>
<tr>
<td>Total</td>
<td>14.87</td>
<td>85.13</td>
<td>100</td>
</tr>
</tbody>
</table>

Furthermore, according to Table 7, we can estimate the precision and recall for the ground-truth datasets, as shown in Table 9. The recall of RSPO is low since here we consider the all the detected plantation locations through 2001 to 2014, and RSPO misses many locations that grow into plantation after 2009. It is noteworthy that here the estimated precision and recall of TP is similar to the assessment result of precision (79%) and recall (94%) provided by Peterson et al. (2016).

Table 9: The precision and recall of ground-truth datasets.

<table>
<thead>
<tr>
<th>Region</th>
<th>precision</th>
<th>recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>TP</td>
<td>0.7320</td>
<td>0.9356</td>
</tr>
<tr>
<td>RSPO</td>
<td>0.9505</td>
<td>0.4700</td>
</tr>
</tbody>
</table>

5.6 Limitation of the method and future work

Before we draw a conclusion, here we present the limitation of our proposed method on two aspects, and introduce our future work on these directions.

First, the performance of the predictive model may be degraded when generalized on a new test area if the training data cannot well cover the possible land cover types in the test area. In our future work, we would like to explore the active learning approach (Tuia et al., 2009) to collect samples from an
iterative process, or investigate the learning approach using imperfect labels (Tabassian et al., 2012), such as the heuristics based on EVI change.

Moreover, we wish to utilize Landsat data which have a superior resolution over MODIS. Although the high resolution (30 m) offers potential to map plantations more accurately, the low temporal frequency (16 day) of Landsat dataset makes it hard to find images with little noise (e.g. clouds). In addition, we need to consider the extra computational cost brought by high-resolution images. In our future work we will explore the learning approach that is more robust to the noise and the multi-scale learning framework that combines Landsat and MODIS.

6. Acknowledgement
This work was supported by the National Science Foundation Award 1029711 and the NASA Award NNX12AP37G. Access to data storage and computing facilities was provided by the University of Minnesota Supercomputing Institute and the NASA Earth Exchange.

7. Conclusion
In this paper we propose an ensemble learning method to map plantation areas. We aggregate the land cover types defined by RSPO and propose to simultaneously sample from multiple land cover types. To learn each individual classifier, we utilize DBN to extract representative information from spectral features over multiple days. Furthermore, we utilize HMM and spatial information to post-process the result. The big remote sensing data of MODIS provide advantage for the learning process, and our test on Kalimantan region of Indonesia (MODIS tile h29v09) shows that 1) the proposed method can generate high-quality annual plantation map, and 2) the detected plantation achieves a better balance of precision and recall than TP and RSPO. The yearly plantation detection can also facilitate the analysis
associated with other natural factors, such as forest fires etc. In addition, the proposed method provides potential to map other specific land cover types with heterogeneous data and generate maps over a regular interval.
References


