Mapping deforestation from palm oil expansion in Sabah, Malaysian Borneo, using SVM classification and change detection

Julia Janicki

Introduction

Deforestation in Borneo has been taking place at a large scale since the 1960s, and since the 1980s and 1990s, it started happening at an unprecedented rate. The palm oil industry has exacerbated it as it started greatly expanding starting from the 2000s, and it is rapidly encroaching on the remnants of primary rainforests in Borneo, and destroying habitats of many endangered species. Deforestation is not only bad as it destroys wildlife habitats leading subsequent species extinctions, it also accounts for about 10% of global CO2 emissions (Van der Werf et al. 2009).

Sabah is a state that is part of Malaysian Borneo, and it particularly experienced rapid increase in palm oil plantations over the past decades. One of the more biodiverse areas in Sabah is forested areas along the Kinabatangan river, and includes endangered species such as the pygmy elephant, Bornean orangutan, and the sunda pangolin (IUCN).

The aim of this project is to assess how much area has been deforested due to palm oil expansion and how it may affect different species. The objectives of this project include:
(1) To classify the 2018 scene at path 117 row 56 to calculate the percent of the land area is palm oil, then intersect the palm oil class of the classified map with ranges of three different species (*Pongo pygmaeus*, *Elephas maximus* and *Neofelis diardi*) in order to get the total deforested areas that coincide with threatened species’ ranges within that scene. (2) To conduct a change detection over two time periods (1997 and 2018) for the Landsat scene at path 117 row 56, a scene that includes the Kinabatangan river area, and conduct an accuracy assessment.

**Background**

The oil palm (*Elaeis guineensis*) is a palm species that is cultivated in many tropical areas across the world, with Malaysia and Indonesia being two major producers. Given that it has the highest oil-yielding capability among oil-bearing crops, oil palm has gained popularity over the years and palm oil has become the most consumed vegetable oil in the world and it accounted for 35% of all vegetable oil consumed in 2016 (Chong et al 2017). A lot of land, in particular forest areas, has been converted to palm oil plantations to satisfy consumer demand, in particularly so in Indonesia and Malaysia. Southeast Asia in general includes 20% of the world’s remaining tropical rainforest, but is among the regions with the highest deforestation rates (Geist & Lambin 2001).

The state of Sabah in particular has experienced rapid increase in palm oil plantations over the past 30-40 years, and it is the state in Malaysia that has the highest planted area of palm oil (Morel, Fisher & Malhi, 2012), accounting for more than 29 percent of national oil palm coverage (MPOB, 2012), producing 10 percent of the world’s annual palm oil output. This leads to great deforestation that in turn leads to many problems such as biodiversity loss. The lower Kinabatangan region in eastern Sabah is particularly well known for its species
diversity and abundance, but due to great deforestation over the past decades there is a decline of many species. For example, the number of Bornean orangutans has fallen from more than 4,000 in the 1960s to 1,125 individuals in 2001, and fewer than 800 individuals in 2017 (Jonas, Abram & Ancrenaz, 2017).

To get an idea of the level of deforestation occurring due to palm oil expansion, satellite remote sensing proves to be a cost-effective, and timely means (Hadi et al. 2018). Over the past several decades, remote sensing data has been widely used for evaluating changes in land-use dynamics as well as for monitoring of degradation (Singh, Malhi & Bhagwata, 2014). Specifically, palm oil has been mapped for different purposes such as detecting illegal deforestation, monitoring spreading of disease, yield estimation, palm counting, and monitoring environmental degradation (Chong et al. 2017). It has also been mapped with different satellites at different scales, including hyperspectral satellites, high resolution satellites such as IKONOS (Thenkabail et al. 2004) and Quickbird (Balasundram, Hadi & Khosla, 2013), medium resolution satellites such as Landsat, low-spatial resolution data such as MODIS, as well as radar images such as PALSAR (Li et al. 2015).

The most relevant studies for this study are listed below. Singh, Malhi and Bhagwata (2014) used remote sensing methods to assess forest stand structure and aboveground biomass in East Sabah, Malaysia. They were able to differentiate old growth forest, logged areas and palm oil plantations by using spectral characteristics as well as texture analysis of Landsat images. Nooni et al. (2014) successfully mapped oil palm related land cover in Ghana by using Maximum Likelihood (ML) and Support Vector Machine (SVM) with Landsat ETM+ images, and concluded that SVM is the preferred method to map palm oil. The study by Morel, Fisher and Malhi (2012) took place in a similar area in Sabah and the authors explored the potential of using Landsat Enhanced ETM+ images to quantify the expansion of palm oil areas and changes
in aboveground biomass in the study area from 2000 to 2008. The study by Hadi et al. (2018) assessed the potential of using Landsat time series data and dense time series algorithm approaches to monitor deforestation in Kalimantan, the Indonesian part of Borneo. Li et al. (2015) explored mapping palm oil in Cameroon using three different classification algorithms (SVM, decision tree and K-means) using PALSAR 50 m Orthorectified Mosaic images.

**Study Area**

The Landsat scene at path 117 and row 56 covers the area in the north eastern part of Sabah, Malaysia. The state of Sabah is situated in northern Borneo and covers 73,965 km² in area. Annual precipitation ranges from 2000 to 3000 mm and annual mean temperature ranges between 26.7–27.7 °C. A lot of deforestation has been happening over the last couple of decades in this area as this area has produced the most palm oil out of all Malaysian states, with oil palm plantations taking up 19.3 percent of Sabah’s land (14,300 km²) already back in 2011 (MPOB, 2012).

This scene also includes the Kinabatangan river, along which are areas that are particularly well known for species diversity and abundance. However, due to great deforestation over the past decades there is a decline of many species in this area.

**Methods**

Two satellite images, one Landsat 5 TM collections 1 level-2 product image (Figure 1.) and one Landsat 8 OLI collections 1 level-2 image (Figure 2.), were acquired from USGS’s Earth Explorer for the years 1997 and 2018 respectively for the scene at path 117 and row 56. Both images had less than 10% cloud cover. Unsupervised classification (K-means) was carried out
to isolate non-relevant features (clouds and shadows) in both the satellite datasets and were both first masked to remove these features.

A Gaussian High Pass filter was first applied to both images in order to enhance local image variations. The kernel size used was 3 x 3 and the image add back was set to 0. A texture analysis was then applied to the red band of both images in order to obtain the following texture measures: mean, variance, homogeneity, contrast, dissimilarity, entropy, second moment and correlation. Texture-based methods can provide valuable insights into the variation of forest structure as can be helpful to differentiate forest from palm (Singh, Malhi & Bhagwata 2014). According to the co-occurrence matrix texture analysis carried out by Gallardo-Cruz et al. (2012), the texture variables derived from the red band are the best for explaining vegetation attributes. The window size used for the texture analysis was 3x3. The filtered image and 8 combination of texture measures for each year were then stacked with the all the bands of the original 1997 and 2018 images respectively to create the final stack of images to be classified (2018 image only) and for conducting change detection (both images).

**Objective 1**

A classification to obtain land cover was first conducted on the 2018 image. The image was first analyzed in order to determine which land cover class types to use to classifying the image, and some adjustments were made along the way. The final land cover class types used are the following: mature palm oil, immature palm oil, forest, and urban.

The classification algorithm used to classify the image is Support Vector Machine, a supervised learning algorithm that produces an optimal hyperplane by determining the location of decision boundaries that produce the optimal separation between classes. Given a set of training samples, the SVM training algorithm search for the hyperplane that minimizes
the training error and maximizes the distance between itself and the closest training samples of each class. Between 15-20 training samples were created for each of the five classes, the radial basis function was used as the kernel type, the gamma in kernel function was set to 0.045 (the default), the penalty parameter that defines the level of misclassification allowed was set to 120, and the classification probability threshold was also set to 0.

An accuracy assessment was then conducted on the classified 2018 image. A random testing sample of 204 pixels (approximately 50 pixels for each class) was created and the testing dataset was then labeled using Google Earth as a reference.

The classified 2018 image was then imported to ArcGIS and vectorized using the Raster to Polygon tool. The area of each land cover class was calculated by using the field calculator. The palm oil class and immature palm oil class of the vectorized image were then intersected with the ranges of the following threatened species in order to obtain the area of their ranges within the scene that have been taken over by palm oil.

Objective 2

The two images (1997 and 2018) were first stacked together and a change detection was conducted on the stacked image to obtain land cover changes between the 1997 and 2018. The images was first analyzed in order to determine which change classes to use for classifying the image, and some adjustments were made along the way. The final change classes produced were: forest to immature palm, forest to palm, forest to urban, immature palm to mature palm, mature palm to immature palm, forest to forest (no change), urban to urban (no change), immature palm to immature palm (no change), mature palm to mature palm (no change), and water to water (no change). Some of the classes were not relevant and were masked while other classes were merged together in order to produce the following more meaningful final change
classes: forest to palm, palm to palm (no change), forest to forest (no change) and urban to
urban (no change). The classification algorithm used to conduct the change detection was also
the SVM algorithm, with the radial basis function used as the kernel type, 0.045 set for the
gamma in kernel function, 100 set for the penalty parameter, and 0 set for the probability
threshold.

An accuracy assessment was then conducted on the change detection image. A
stratified random testing sample of 202 pixels (approximately 50 pixels for each class) was
created and the testing dataset was then labeled using Google Earth and the optical images of
1997 and 2018 as references.

Results

The result of the SVM classified image of 2018 indicates that there are a lot of palm oil
plantations in the study area (Figure 3.). The land area of scene 117, 56, after cloud areas were
removed is a total of 26198.15 km² (5990.7 km² mature palm oil, 1844.67 km² immature palm
oil, 17226 km² forest, 1136.78 km² urban). The total palm oil (including mature and immature
palm) is 7835.37 km², which is 29.9% of the land area in this scene.

The area of different species ranges in this scene that coincides with palm oil are:
2987.67 km² for the Bornean orangutan (*Pongo pygmaeus*) (Figure 5.), 1140.52 km² for the
pygmy elephant (*Elephas maximus*) (Figure 6.), and 3190.91 km² for the Bornean clouded
leopard (*Neofelis diardi*) (Figure 7.).

The overall accuracy of the SVM classified image of 2018 is 71.5686% (146/204) and the
Kappa coefficient is 0.5263 (Table 1.). The forest class has the highest user’s accuracy at
77.36%, which means that users would find that most of the time when they visit this area that
what the classified map calls “forest” will actually be forest, most other times it would actually
be mature palm. It also has the highest producer’s accuracy at 80.39%, which means that most of the forest areas have been correctly identified as “forest”, some areas that should have been classified as forest were misclassified as mature palm. “Immature palm” has a 62.5% user’s and producer’s accuracies, and often what actually should be “urban” are misclassified as “immature palm”, some areas that should be classified as “immature palm” were misclassified as “mature palm”. “Mature palm” has a 66.67% user’s accuracy and 63.16% producer’s accuracy, and often what should be “forest” are misclassified as “mature palm”, some areas that should be classified as “mature palm” were misclassified as “forest”. “Forest” and “mature palm” are the most confused classes, which is logical since mature palm oil and forest are both dense vegetation.

The result of the change detection shows that there are a lot of forested areas that has been converted to palm oil between the years 1997 and 2018 (Figure 4.).

The overall accuracy of the change detection image is 68.8119% (139/202) and the Kappa coefficient is 0.5791 (Table 2.). “Palm to palm” has the highest user’s accuracy at 77.59%, which means that users would find that most of the time when they visit this area that what the classified map calls “palm” will actually be either mature or immature palm oil areas both in 1997 and 2018. However, the producer’s accuracy of the same class is only 59.21%, which means that some areas that should be forest or urban areas in both 1997 and 2018 were misclassified as palm areas in both time periods. The “forest to palm” class has a 63.64% user’s accuracy, which means that only 63.64% of the time that an area that has changed from forest to palm has actually gone through that change, other times they were simply forest areas or palm areas. It has a 71.43% producer’s accuracy, which means that some areas that should have been classified as “forest to palm” were in fact forest, palm or urban areas in both time periods. Again, forest and mature palm share similar spectral characteristics given they are
both dense vegetation, which is why “palm to palm”, “forest to forest” and “forest to palm” are classes that are most often confused.

**Discussion**

The overall accuracy of the SVM classification of the 2018 image and change detection are in general satisfactory, but further improvements could be made.

The overall accuracy is 71.5686% for the 2018 classified image and 68.8119% for the change detection image, which is lower compared to 78.29% in the study to map palm oil by Nooni et al. (2014) that also used SVM to classify palm oil areas. However, the results are about similar in accuracy compared to the 69.7% accuracy of the classified Landsat image in the study by Morel, Fisher and Malhi (2012), but lower compared to the 97.0% accuracy of the classified ALOS-PALSAR image in the same study.

The study by Morel, Fisher and Malhi (2012) was similarly mapping the increase in palm oil areas in Sabah, but across a greater extent and from 2000 to 2008. The study found a 38% increase (1450 km2) in oil palm areas and a 13.1% total decrease (1900 km2) in forest areas among the whole study area within the eight year period. According to this study, as well as results indicated by the accuracy assessment of the 2018 classification and change detection, it seems that the palm oil areas in the present study are underestimated and there should be more forest areas that were converted to palm oil.

SVM was the algorithm of choice for both analyses as supervised learning algorithms tend to be more accurate than unsupervised learning algorithms, and in various studies (Nooni et al. 2014) SVM seems to be the algorithm of choice for mapping palm oil, in particular it is more accurate than Maximum Likelihood or K-means. However, different classification
algorithms could be explored for mapping palm oil to potentially improve the results. According to Li et al. (2015), though SVM performed better in general, for large-scale mapping of palm oil plantations the Decision Tree algorithm outperformed both SVM and K-Means in terms of speed and performance.

An option to improve the results is to use vegetation indices (VIs). Normalized Difference Vegetation Index (NDVI) is among the most widely used VIs in remote sensing due to its reduction of topographic effects and its normalized linear relationship between the R and NIR bands, but it has not been found to be a reliable VI for mapping areas of dense biomass such as tropical forests (Foody et al. 2003). However, other VIs can be used to map palm oil, such as the Soil-Adjusted Vegetation Index (SAVI) and the Normalized Difference Infrared Index (NDII) used by Singh, Malhi & Bhagwata (2014), or the Normalized Difference Moisture Index (NDMI) which was observed by Hadi et al. (2018) as the VI with the highest sensitivity in response to deforestation events.

Moreover, specific bands could be used instead of all bands, such as band 4, band 5 and band 3 for Landsat ETM+ as suggested by Nooni et al. (2014). Bands 4, 5, 3 contains a near infrared red (band 4), mid infrared red (band 5) and red(band 3) which can be clearly defined for different vegetation types because they exhibit variations in moisture.

Finally, another option would be to use different types of imagery such as high spatial resolution imagery or radar imagery. Li et al. (2015) and Morel, Fisher and Malhi (2012) achieved results with higher accuracy by using radar images. Moreover, high resolution images can be used to achieve better results since they can provide more details of the palm oil areas.

**Conclusion**
After classifying the area covered by the Landsat scene at path 117 row 56 in Sabah, Malaysia in 2018 and conducting a change detection of the same scene between 1997 and 2018 by using SVM classification, the results indicated that there is a great increase in palm oil areas from 1997 to 2018, many of which used to be forested areas. The areas classified as palm oil also had great overlaps with ranges of threatened species such as the Bornean orangutan, the Asian elephant and the Bornean clouded leopard.

The results were satisfactory with an overall accuracy of 71.5686% for the 2018 classified image and 68.8119% for the change detection image, with mature palm and forest areas the most often confused. The results could be improved by using vegetation indices, using different band combinations, using different types of imagery such as high spatial resolution optical data or radar data, using a different classification algorithm, or refining the training dataset.

Reference


Figures

**Figure 1.** The true color image (band 3, band 2, band 1) of an area in Sabah, Malaysia at path 117 and row 56. The acquisition date of this image is 1997 August 10, the resolution of the image is 30 m, and the image was acquired by Landsat 5 TM with bands 1 to 5 plus 7.
**Figure 2.** The true color image (band 4, band 3, band 2) of an area in Sabah, Malaysia at path 117 and row 56. The acquisition date of this image is 2018 August 20, the resolution of the image is 30 m, and the image was acquired by Landsat 8 OLI, with bands 1 to 7 plus 9.
Figure 3. The final classified map of the 2018 image that displays different land cover classes (forest, mature palm, immature palm and urban) by using the SVM supervised learning algorithm in an area in Sabah, Malaysia at path 117 and row 56. The acquisition date of this image is 2018 August 20, the resolution of the image is 30 m, and the image was acquired by Landsat 8 OLI, with bands 1 to 7 plus 9.
Figure 4. The final change detection map of the 1997 and 2018 images that displays different land cover class changes (forest to forest, forest to palm, palm to palm and urban to urban) between the two periods by using the SVM supervised learning algorithm in an area in Sabah, Malaysia at path 117 and row 56. The acquisition date of the two images are 1997 August 10 and 2018 August 20, the resolution of the image is 30 m, and the 1997 image was acquired by Landsat 5 TM with bands 1 to 5 plus 7. while the 2018 image was acquired by Landsat 8 OLI, with bands 1 to 7 plus 9.
Figure 5. This map displays the areas in the range of the Bornean orangutan (*Pongo pygmaeus*) within the Landsat scene at path 117 and row 56 that overlap with palm oil plantations.
**Figure 6.** This map displays the areas in the range of the Asian elephant (*Elephas maximus*) within the Landsat scene at path 117 and row 56 that overlap with palm oil plantations.
Figure 7. This map displays the areas in the range of Bornean clouded leopard (*Neofelis diardi*) within the Landsat scene at path 117 and row 56 that overlap with palm oil plantations.
Tables

### Table 1

<table>
<thead>
<tr>
<th>Class</th>
<th>immature palm test</th>
<th>urban test</th>
<th>forest test</th>
<th>mature palm test</th>
<th>Total</th>
<th>User's Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>immature palm</td>
<td>10</td>
<td>4</td>
<td>0</td>
<td>2</td>
<td>16</td>
<td>62.5%</td>
</tr>
<tr>
<td>urban</td>
<td>2</td>
<td>6</td>
<td>0</td>
<td>2</td>
<td>10</td>
<td>60%</td>
</tr>
<tr>
<td>forest</td>
<td>0</td>
<td>0</td>
<td>82</td>
<td>24</td>
<td>106</td>
<td>77.36%</td>
</tr>
<tr>
<td>mature palm</td>
<td>4</td>
<td>0</td>
<td>20</td>
<td>48</td>
<td>72</td>
<td>66.67%</td>
</tr>
<tr>
<td>Total</td>
<td>16</td>
<td>10</td>
<td>102</td>
<td>76</td>
<td>204</td>
<td></td>
</tr>
</tbody>
</table>

**Producer's Accuracy**

<table>
<thead>
<tr>
<th>User's Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>62.5%</td>
</tr>
<tr>
<td>60%</td>
</tr>
<tr>
<td>80.39%</td>
</tr>
<tr>
<td>63.16%</td>
</tr>
</tbody>
</table>

**Table 1.** Accuracy assessment of the SVM classification of the 2018 image, with overall accuracy being 71.5686% (146/204) and the Kappa coefficient being 0.5263.

### Table 2

<table>
<thead>
<tr>
<th>Class</th>
<th>Forest to palm test</th>
<th>Forest to forest test</th>
<th>Urban to urban test</th>
<th>Palm to palm test</th>
<th>Total</th>
<th>User's Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forest to palm</td>
<td>35</td>
<td>12</td>
<td>0</td>
<td>8</td>
<td>55</td>
<td>63.64%</td>
</tr>
<tr>
<td>Forest to forest</td>
<td>5</td>
<td>32</td>
<td>0</td>
<td>11</td>
<td>48</td>
<td>66.67%</td>
</tr>
<tr>
<td>Urban to urban</td>
<td>2</td>
<td>0</td>
<td>27</td>
<td>11</td>
<td>40</td>
<td>67.5%</td>
</tr>
<tr>
<td>Palm to palm</td>
<td>7</td>
<td>4</td>
<td>2</td>
<td>45</td>
<td>58</td>
<td>77.59%</td>
</tr>
<tr>
<td>Total</td>
<td>49</td>
<td>48</td>
<td>29</td>
<td>76</td>
<td>202</td>
<td></td>
</tr>
</tbody>
</table>

**Producer's Accuracy**

<table>
<thead>
<tr>
<th>User's Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>71.43%</td>
</tr>
<tr>
<td>66.67%</td>
</tr>
<tr>
<td>93.1%</td>
</tr>
<tr>
<td>59.21%</td>
</tr>
</tbody>
</table>

**Table 2.** Accuracy assessment of the change detection image, with the overall accuracy of the being 68.8119% (139/202) and the Kappa coefficient being 0.5791.