DEVELOPMENT OF A NEW ALLOMETRIC EQUATION CORRELATED WITH RS VARIABLES FOR THE ASSESSMENT OF DATE PALM BIOMASS

Salem Issa (1); Basam Dahy (1); Taoufik Ksiksi (1), and Nazmi Saleous (2)

1 UAE University, College of Science, P.O. Box 15551, Al Ain, United Arab Emirates
2 UAE University, College of Humanities and Social Sciences, P.O. Box 15551, Al Ain, United Arab Emirates

Email: salem.essa@uaeu.ac.ae; basam.d@uaeu.ac.ae; tksiksi@uaeu.ac.ae; nazmi.saleous@uaeu.ac.ae

KEY WORDS: Arid lands; Biomass; Carbon Sequestration; Remote Sensing; UAE

ABSTRACT: The United Arab Emirates (UAE) has the largest number of date palms for any single country in the world with an estimated number of about 42 million palms. Date palm, Phoenix dactylifera, is considered as the most important fruit crop in arid regions. In addition to its ability to tolerate harsh weather, high temperature, drought and high levels of salinity, date palm is a good source for carbon sequestration in such ecosystems. The main objectives of the actual study were to develop new allometric equations for aboveground biomass assessment of date palm in the arid lands of the UAE and to determine the best remote sensing (RS) predictors to develop a RS-based biomass assessment model. Testing plots from Al Ain area (UAE) were used to estimate the above ground biomass (AGB) of date palm using standard sampling techniques. It was found that crown area (CA) and trunk height (Ht) of the date palm were the best field measured variables for the prediction of AGB. In each plot, crown area (CA) and trunk height (Ht) of date palms were measured to estimate the biomass at plot level. These findings were tested for their correlation with remote sensing (RS) variables (derived from Landsat 8 bands) to determine the best predictors and develop a RS based biomass assessment model. The relationships between the AGB and RS variables (e.g., single bands and various vegetation indices) were investigated using multiple linear regression analysis. Preliminary results were promising, and final model is being built and will be published soon.

1. INTRODUCTION

Many techniques exist to assess biomass and estimate carbon stock in forests (Gibbs, Brown, Niles, & Foley, 2007). All techniques ultimately rely on ground measurement of tree biomass which is time consuming, tedious and destructive. These reasons justify the development of alternative methods for forest biomass estimation as the destructive methods cannot be used as a routine procedure. Allometric equations are developed to be used to predict tree biomass from dendrometric measures, such as tree diameter and tree height, and to estimate carbon stocks (Ebuy, Lokombe, Ponette, Sonwa, & Picard, 2011; Picard, Saint-André, & Henry, 2012). The number of trees destructively sampled to build allometric equations differ from one researcher to another: (Russell, 1983) used 15 trees; (Deans, Moran, & Grace, 1996) used 14; (Khalid, Zin, & Anderson, 1999a) used 10 trees; while (I. F. Brown et al., 1995) used only a sample of 8 trees.

Most of biomass equations, species and multi-species, have been developed for tropical rainforests ecosystems because of their importance to the global carbon cycle (S. Brown, 1997; Chave et al., 2005). Unfortunately, very few plant species biomass measurements are available for desert ecosystems. None of these measurements were used to fit one of the most important fruit crops in the arid regions, Phoenix dactylifera, date palm (Chao & Krueger, 2007). According to Food and Agriculture Statistics Database (FAOSTAT, 2013) the total world number of date palms is about 120 million trees, distributed in 30 countries and producing nearly 7.5 million tons of fruit per year. Over two-third of this amount of date palms are in Arab countries (El-Juhany, 2010). The United Arab Emirates (UAE) has the largest number of date palms for any single country in the world. In 2006, the UAE had 16,342,190 productive date palms producing 757,600 tons of dates. Furthermore, it was proclaimed on 15 March 2009 that the UAE had reached the 42 million date palm trees (El-Juhany, 2010) with a minimum of 200 cultivars, 68 of which are the most important commercially.

The objectives of the actual research are: (1) to develop a specific allometric biomass estimation equation for date palm in the UAE; (2) to find best remote sensing variables that can be used to predict biomass in date palms and; (3) to correlate these RS predictors with field measurements and allometric equations.
2. STUDY AREA

The study area is located in Al-Ain area, the capital of the eastern district of Abu Dhabi emirate, the capital of the United Arab Emirates (UAE). Al-Ain city was established around an old date palm oasis (*Phoenix dactylifera* L) and as the city expanded, the date palm plantations expanded as well. The estimated number of date palms in Al-Ain area is 8.5 M palms which represent more than fifth of the total date palms of the whole country (“Date Palm Culture in UAE,” n.d.). The minimum average temperature in Al-Ain is 22°C while the average maximum is 35.8°C. The annual average long-term rainfall is 119.7 mm and the humidity is 58%. Three private farms in Al-Ain area are selected to conduct the field data collection. Namely: Masakin, Qattara and Salamat West.

![Map of Al Ain City, UAE](image)

*Note: Blue points represent the farms where the fifteen date palm were sampled. The image is a subset of scene no (160-43) - Landsat 8 (17th June, 2017). It is a false color composite image. Red color is representing vegetation. Figure 1. Al Ain City, UAE.*

3. METHODS

3.1. Field Data Collection

Date palms are selected to run the destructive method to estimate aboveground biomass (AGB), below ground biomass (BGB), building allometric equation specifically for date palm and calculating the carbon stored in biomass for date palm plantations in the Emirate of Abu Dhabi, where date palm plantation is the most dominant agricultural activity. Age is one of the most important factors that influence the biomass of the palm and its structural measurements (Sunaryathy, Kanniah, & Tan, 2015). A lot of work has been done on the estimation of oil palm biomass at various ages (Khalid, Zin, & Anderson, 1999). Based on, three age classes of date palms are targeted: Class 1, covering plantations with age < 5 years; class 2, with age between 5 and 10 years; and class 3, covering matured forests exceeding 10 years age. Five date palms were selected from each age class.

Another influencer factor in date palm biomass storing is variety. For, date palms in the study area differ in their cultivars (varieties) as well, which has its own effect on palm growth rate. We tried to find 15 date palms of different cultivars (varieties) to represent the whole date palm cultivars in the study area and in the country as much as possible. The date palm cultivars found in the study area are: Fardh, Bumaan, Khunaizi, Khlalas, Baghel, Jabri, Shahem, Jash Ramli and Neghal.

3.2. Field Measurements

A fieldwork was conducted during 22nd -29th April 2018 (including the initial visits to farms for palms selection). Five date palm trees were uprooted for each age class (total of 15 trees). These uprooted trees were partitioned into three parts: crown, trunk and roots (Khalid et al., 1999). Above ground biomass (AGB) is defined in this research as the sum of the crown and trunk weight while below ground biomass (BGB) is defined as the weight of the root.
A big scale balance was used to get the fresh weight of crown, trunk and roots in kg. From each part of the uprooted palms, three samples were collected (3 crown samples, 3 trunk samples and 3 root samples), (Figure 2).

Figure 2: Uprooting date palms, partitioning the palm: crown, trunk and roots, and weighing the palm's parts.

Structural variables of date palm (total palm height, trunk height, diameter breast height, crown diameter, crown area and number of fronds) were measured to be used later in the multi-regression analysis and for building up specific biomass allometric equations of date palm in the UAE that relate field variables and biomass through different permutations and combinations (versus crown biomass, trunk biomass, root biomass, AGB and total biomass).

Before uprooting the palms, the following variables were measured: (a) Diameter at breast height (DBH) in cm by measuring the circumferences of the trunk at 1.3 m height and dividing by \( \pi \). For the small palms where there is no trunk developed yet, the diameters were measured at the base of the palm, (b) Number of palm fronds (\#Frond), (c) Crown diameter (CD) in meter, and (d) Crown area (CA) was calculated using sphere equation (\( CA = \pi CD^2/4 \)), assuming a rounded palm crown. After uprooting the palms, the following variables were measured (Ebuy et al., 2011): (a) Palm height (H) in meter, (b) Palm trunk height (Ht) in meter, and (c) Crown depth (a difference between total and trunk (\( \Delta \)height) in the meter.

3.3. Sample Processing And Biomass Calculation

A total of 120 samples were collected during the fieldwork [(15 Crown + 10 Trunk + 15 Root) X 3 replicates] (Note that small palms do not have developed trunk). All samples were stored in plastic bags and then transported to the UAE University labs for further processing and data analyses to assess the percent organic matter (OM) and organic carbon (OC) in each sample (Khalid et al., 1999).

3.4. Palm Samples’ Biomass Calculation

Biomass samples were brought to the lab and were immediately weighted at arrival at the lab to measure the fresh Weight (g) of the sample. Next, samples were air dried and transferred to paper bags to be ready for being oven-dried at 80°C for 72-96 hours to get rid of the water content (Khalid et al., 1999). Then, samples were weighted to get the percentage of dry weight to original fresh weight in each sample (dry to fresh factor=DF) (Figure 3). The following calculations were done:

i. Calculate the Dry Weight of each palm part:

\[
\text{Crown Dry Weight (kg)} = \text{Crown Fresh Weight (kg)} \times \text{Crown DF} \quad (1) \\
\text{Trunk Dry Weight(kg)} = \text{Trunk Fresh Weight (kg)} \times \text{Trunk DF} \quad (2) \\
\text{Root Dry Weight (kg)} = \text{Root Fresh Weight (kg)} \times \text{Root DF} \quad (3)
\]

ii. Calculate the percentage of BGB (Root system) from the AGB and other ratios.
iii. Calculate the AGB weight in each palm

\[ AGB \text{ (kg)} = \text{Crown Dry Weight} + \text{Trunk Dry Weight} \]  

(4)

iv. Calculate the Total Biomass of each palm

\[ \text{Total Biomass (kg)} = \text{AGB Weight} + \text{Root biomass Weight (BGB)} \]  

(5)

Figure 3: Lab works: Preparing samples, weighing, and drying

The correlation coefficient for each date palm part (crown, trunk and root) with all collected field variables was calculated. Then, a regression analysis was performed using different types of regressions (linear, logarithmic, exponential, power and polynomial regressions) to find the highest coefficient of determination \((r^2)\). Scatter plots were drawn to visualize the relationship between crown biomass versus the field variables to identify the most important predictors for biomass estimation for each date palm part.

4. RESULTS

4.1. Measurements Of Biomass

In this study, a total of 15 stands of date palm were sampled and classified in three different age classes: young, medium and mature. Their characteristics in each one of these classes are summarized in Table 1. Note that the crown area and crown depth (Δ height) are calculated as shown in section # 3.2. above.

It is shown that the dry weight is highly correlated with fresh weight for all three parts: the crown, trunk and root (0.991, 0.977 and 0.978; respectively). This high correlation is usually reported in similar palm experiments (R. H. V. Corley, Hardon, & Tan, 1971). The correlation between the total fresh weight of date palm and dry weight was 0.994. The dry weights to fresh weights factor (DF) in belowground (root system) ranged from 0.32 to 0.59 with mean equal 0.45 which is higher than DF of aboveground (crown plus trunk), which ranged from 0.31 to 0.51 with mean of 0.40.

Age is an important factor influencing the storing of date palm biomass. This relationship between age of date palm and its biomass remains strong and positive in the case of fresh weight or dry weight and for the total weight or for only the aboveground weight. (Table 1).

Table 1: Characteristics of date palms at different age classes and the mean fresh and dry weight of each part.

<table>
<thead>
<tr>
<th>Component</th>
<th>Young (&gt;5 yrs) (Mean ± S.E)</th>
<th>Med. (≥5 yrs, ≤10 yrs) (Mean ± S.E)</th>
<th>Mature (&gt;10 yrs) (Mean ± S.E)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DBH (cm)</td>
<td>(33.87 ± 2.28)</td>
<td>(43.29 ± 7.45)</td>
<td>(51.57 ± 5.1)</td>
</tr>
<tr>
<td>H (m)</td>
<td>(4 ± 0.167)</td>
<td>(4.85 ± 0.206)</td>
<td>(8.38 ± 0.48)</td>
</tr>
<tr>
<td>Ht (m)</td>
<td>-</td>
<td>(0.764 ± 0.196)</td>
<td>(3.21 ± 0.52)</td>
</tr>
<tr>
<td>Δ height (m)</td>
<td>(4 ± 0.17)</td>
<td>(4.086 ± 0.22)</td>
<td>(5.17 ± 0.36)</td>
</tr>
<tr>
<td>CD (m)</td>
<td>(3.09 ± 0.46)</td>
<td>(5.66 ± 0.25)</td>
<td>(7.2 ± 0.08)</td>
</tr>
<tr>
<td>CA (m²)</td>
<td>(8.15 ± 2.57)</td>
<td>(25.36 ± 2.28)</td>
<td>(40.73 ± 0.86)</td>
</tr>
<tr>
<td># Frond</td>
<td>(29.8 ± 2.27)</td>
<td>(35 ± 5.17)</td>
<td>(61.6 ± 2.32)</td>
</tr>
</tbody>
</table>
Wt. of fresh parts (kg)

1) Crown (50.65 ± 5.43) (171.08 ± 34.47) (367.24 ± 78.56)
2) Trunk - (74.18 ± 13.61) (365.28 ± 30.65)
3) Root (21.43 ± 6.39) (187.36 ± 27.91) (282.06 ± 25.25)
Total Wt. (kg/palm) (72.08 ± 11.19) (432.62 ± 66.41) (1014.58 ± 95.92)
Aboveground Wt. (kg/palm)* (50.65 ± 5.43) (245.26 ± 42.99) (732.52 ± 91.38)

Wt. of dry parts (kg)

1) Crown (22.51 ± 3.06) (65.17 ± 11.87) (148.5 ± 35.85)
2) Trunk - (29.53 ± 8.62) (135.91 ± 19.62)
3) Root (7.46 ± 1.88) (87.61 ± 14.87) (141.23 ± 13.59)
Total Wt. (kg/palm) (29.97 ± 4.17) (182.3 ± 32.07) (425.63 ± 45.6)
Aboveground Wt. (kg/palm)* (22.51 ± 3.06) (94.69 ± 18.45) (284.41 ± 43.15)

Where DBH is diameter breast height in m, H is total height in m, Ht is trunk height in m, Δ height is the difference between total height and trunk height in m or referred to as crown depth, CD is a crown diameter in m, CA is the crown area in m² and # Frond is the number of fronds. *Aboveground weight is equal crown weight plus trunk weight of the palm.

4.2. Biomass Ratios

The correlation between fresh weight and dry weight of date palm was very high (0.99). Therefore, dry weight was considered in all later calculations and for building the biomass allometric equations for date palm. Wherever the term biomass is present in this paper it means dry weight except when indicated otherwise. It is worth indicating here that some researchers prefer to use fresh weight instead of dry weight for building their equations (Dewi, Khasanah, Rahayu, Ekadinata, & Van Noordwijk, 2009; Khalid et al., 1999).

In young palms, with no main trunk, the crown biomass ranged between 17.26 kg and 34.19 kg with the mean value of 22.51 kg, which contributed to 75.11% of the total palm biomass (Table 2). The crown biomass is equal to the aboveground biomass (AGB) as there is no trunk developed yet at such a early age, while the root system (the belowground biomass (BGB)), contributed to about 24.89% of the total biomass (Table 2). The contribution percentages of the crown and root biomass to the total biomass were dramatically changing with the growing up of the palm and the development of the trunk. The percentage of crown biomass to the total biomass decreases to reach 35.75% and 34.89% of the total biomass in medium and mature palms respectively. Likewise, the root systems were changing in their contribution percentages of the total biomass (Table 2).

Table 2: The ratios of each biomass part to the total biomass and to the above ground biomass at different age classes

<table>
<thead>
<tr>
<th>Total Biomass</th>
<th>Young (&gt;5 yrs)</th>
<th>Med. (≥5 yrs, ≤10 yrs)</th>
<th>Mature (&gt;10 yrs)</th>
<th>(Mean ± S.E)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>% Crown Biomass</td>
<td>75.11</td>
<td>35.75</td>
<td>34.89</td>
<td>(35.32 ± 0.43)</td>
</tr>
<tr>
<td>% Trunk Biomass</td>
<td>-</td>
<td>16.2</td>
<td>31.93</td>
<td>(24.065 ± 7.865)</td>
</tr>
<tr>
<td>% Root Biomass</td>
<td>24.89</td>
<td>48.06</td>
<td>33.18</td>
<td>(40.62 ± 7.44)</td>
</tr>
<tr>
<td>% AGB</td>
<td>75.11</td>
<td>51.94</td>
<td>66.82</td>
<td>(59.38 ± 7.44)</td>
</tr>
</tbody>
</table>

AGB is above ground biomass. BGB is below ground biomass (root systems). * The mean was calculated for only the medium and mature age classes which have already developed a trunk.

4.3. Allometric Equations:

Crown Biomass: All field variables showed a positive correlation with crown biomass except DBH. Age, total height, trunk height, fronds number, crown area and crown diameter showed highly positive correlation (0.88, 0.88,
0.87, 0.81, 0.78 and 0.74; respectively) while the difference between total height and trunk height (Δ height) showed intermediate correlation with crown biomass (0.59). Age as an indicator to crown biomass showed the best coefficient of determination ($R^2 = 0.8323, P < 0.05$) with an exponential regression equation. Crown area, crown diameter and total height were found to have best $R^2$ coefficient value of the seven indicators used in Table 3 and figure 4 below.

Table 3: Regression equations for crown biomass estimation for date palms of Abu Dhabi

<table>
<thead>
<tr>
<th>Regression Equations</th>
<th>Indicator (x)</th>
<th>Correlation</th>
<th>$R^2$</th>
<th>$\rho$ - Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crown Biomass = 16.61e$^{0.1188x}$ = 16.61 * 1.1453$x$</td>
<td>Age</td>
<td>0.879</td>
<td>0.8323</td>
<td>0.000001</td>
</tr>
<tr>
<td>Crown Biomass = 0.2506x$^{-0.0548}$</td>
<td>DBH</td>
<td>0.331</td>
<td>0.3054</td>
<td>&gt; 0.05</td>
</tr>
<tr>
<td>Crown Biomass = 1.0874x$^{-5.3225}$</td>
<td>H</td>
<td>0.879</td>
<td>0.8114</td>
<td>0.000001</td>
</tr>
<tr>
<td>Crown Biomass = 39.191x + 26.81</td>
<td>Ht</td>
<td>0.869</td>
<td>0.7566</td>
<td>0.000002</td>
</tr>
<tr>
<td>Crown Biomass = 0.3013x$^{1.5402}$</td>
<td>Δ Height</td>
<td>0.591</td>
<td>0.4466</td>
<td>0.02</td>
</tr>
<tr>
<td>Crown Biomass = 5.8364e$^{0.8254x}$ = 5.8364 * 1.5267$x$</td>
<td>CD</td>
<td>0.741</td>
<td>0.8143</td>
<td>0.001</td>
</tr>
<tr>
<td>Crown Biomass = 14.034e$^{1.0554x}$ = 14.034 * 1.057$x$</td>
<td>CA</td>
<td>0.780</td>
<td>0.8354</td>
<td>0.0006</td>
</tr>
<tr>
<td>Crown Biomass = 0.1113x$^{-0.64461x} + 125.63$</td>
<td>#Frond</td>
<td>0.807</td>
<td>0.7181</td>
<td>0.0002</td>
</tr>
</tbody>
</table>

Where DBh is diameter breast height in cm. H is the total height in m. Ht is trunk height in m. Δ Height is the difference between total height and trunk height (called crown depth as well). CD is crown diameter in m. CA is crown area in m². #Frond is frond number per palm.

Figure 4: Crown biomass values in function of: (a) crown area; (b) crown diameter and (c) total height.

Trunk Biomass: All field variables showed a positive correlation with trunk biomass. Trunk height, total height, age, fronds number, crown area and crown diameter showed highly positive correlation (0.93, 0.92, 0.91, 0.84, 0.82 and 0.78; respectively) while DBH and the difference between total height and trunk height (Δ height) showed intermediate correlation with trunk biomass (0.6 and 0.67; respectively). Trunk height as an indicator to trunk biomass showed the best coefficient of determination ($R^2 = 0.889, P < 0.05$) with polynomial regression equation (2nd order) (Table 4 and Figure 5).

Table 4: Regression equations for trunk biomass estimation for date palms of Abu Dhabi (UAE).

<table>
<thead>
<tr>
<th>Regression Equations</th>
<th>Indicator (x)</th>
<th>Correlation</th>
<th>$R^2$</th>
<th>$\rho$ - Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trunk Biomass = 10.138x - 34.742</td>
<td>Age</td>
<td>0.908</td>
<td>0.8238</td>
<td>0.000002</td>
</tr>
<tr>
<td>Trunk Biomass = 2.9369x - 70.87</td>
<td>DBH</td>
<td>0.600</td>
<td>0.3604</td>
<td>0.01</td>
</tr>
<tr>
<td>Trunk Biomass = 29.12x - 112.1</td>
<td>H</td>
<td>0.919</td>
<td>0.8443</td>
<td>0.000001</td>
</tr>
<tr>
<td>Trunk Biomass = -3.956 * (Ht)$^2$ + 55.247 * (Ht) - 2.0342</td>
<td>Ht</td>
<td>0.934</td>
<td>0.8838</td>
<td>0.0000003</td>
</tr>
<tr>
<td>Trunk Biomass = 232.15ln(x) - 286.73</td>
<td>Δ Height</td>
<td>0.566</td>
<td>0.3408</td>
<td>0.02</td>
</tr>
<tr>
<td>Trunk Biomass = 27.304x - 90.019</td>
<td>CD</td>
<td>0.776</td>
<td>0.6018</td>
<td>0.0007</td>
</tr>
<tr>
<td>Trunk Biomass = 0.1188x$^{-2.18443x} + 4.9609$</td>
<td>CA</td>
<td>0.822</td>
<td>0.751</td>
<td>0.0002</td>
</tr>
<tr>
<td>Trunk Biomass = 3.4128x - 88.645</td>
<td>#Frond</td>
<td>0.840</td>
<td>0.706</td>
<td>0.00009</td>
</tr>
</tbody>
</table>

Where DBh is diameter breast height in cm. H is the total height in m. Ht is trunk height in m. Δ Height is the difference between total height and trunk height (called crown depth as well). CD is crown diameter in m. CA is crown area in m². #Frond is frond number per palm.
AGB: As the AGB can be calculated as a sum of crown biomass and trunk biomass and as shown above that crown area and trunk height were the most important field structural variable for predicting the crown and trunk (respectively), the AGB can be estimated according to the following allometric equation (Tables 4 & 5):

\[
AGB = \text{Crown Biomass} + \text{Trunk Biomass}
\]

where;

\[
\text{Crown Biomass} = 14.034 \times 1.057^{ca}
\]

and,

\[
\text{Trunk Biomass}^a = -3.956 \times (Ht)^2 + 55.247 \times (Ht) - 2.0342
\]

\(a\) (it is to be noted that for young palms the value of this equation is equal to zero as there is no developed trunk for such age class. Therefore, the amount of AGB for young palms will equal crown biomass alone).

\(ca\) is crown area in square meter and \(Ht\) is trunk height in meter.

#### 4.4. Correlation With RS Variables

Till this stage of the research, the AGB of date palm is the dependent variable of the regression equations. Next step was conducted on plots level where the AGB of date palm plantations was the independent variable of the regression equation and the dependent variables were the remote sensing (RS) variables. By using Landsat 8 bands and vegetation indices (VIs’s), many RS variables were tested for their ability to predict biomass of date palm and to determine their relationships with AGB using field plots data collected from the date palm plantations of Abu Dhabi (Figure 6). The mean values from 3X3 windows over the plots for each of the spectral variables were extracted (Table 5). Landsat bands or VIs alone were not sufficient to establish effective AGB estimates so a combination of these RS variables was used that can be grouped into two distinct categories:

- **Raw Landsat 8 bands:** band1 (coastal/aerosol-B1), band2 (blue-B), band3 (green-G), band4 (red-R), band5 (near infrared-NIR), band6 (shortwave infrared-SWIR1), and band7 (shortwave infrared 2-SWIR).
- **VIs that are characterized by their sensitivity for vegetation biophysical parameters and widely used in RS change detection and biomass estimation.**

Traditional as well as a variety of modified VIs were tested, this includes: simple ratio (SR), ratio vegetation index (RVI), difference vegetation index (DVI), normalized difference index (NDI), normalized difference vegetation index (NDVI), transformed vegetation index (TVI), green normalized difference vegetation index (GNDVI), soil-adjusted vegetation index (SAVI) and the three tasseled cap transformation indices for greenness (TCG), brightness (TCB) and wetness (TCW).

Figure 6: The location of the study area plots in Abu Dhabi.
Vegetation indices such as: NDVI, modified NDVI, GNDVI and NDI have minimum and maximum values range between -1 and +1 where green surfaces are represented by DN values ranging between 0.2 and 0.9 with dense vegetation and higher biomass amount are generally represented by higher DN values (more than 0.6). However, TVI, which is another modified NDVI, has a range of values between 0 and 1.4 with no negative values. On the other hand, positive DN values of DVI and SR represents vegetation with higher values depicting higher biomass amount. Likewise, RVI shows higher values for thick vegetation than for sparse or non-vegetated surfaces. Finally, SAVI index behaves similar to NDVI, ranging between -1 and 1 with lower values reflecting lower biomass amount/cover of green vegetation.

4. DISCUSSION

Fresh weight and dry weight of date palm are highly correlated. Some researchers used fresh weight to build their allometric equations as done in southern Asian oil palms (Khalid et al., 1999), while others used the dry weight as done in tropical regions (R. Hereward V. Corley & Tinker, 2008). In the actual study, we used dry weight to calculate AGB allometric equations of Abu Dhabi date palm. Furthermore, it is recognized that most of allometric equations were built to calculate the AGB while BGB is determined as a ratio of AGB.

Date palm trunks were found to contain between 22% to 51% of dry matter from fresh material? with an average value of 37%, in comparison to oil palm trunks which contain only about 20% (Khalid et al., 1999). We also found that AGB of the dry weight of date palm is estimated to have a ratio of 40% of the AGB of fresh weight. In comparison, the published constant converter from fresh to dry AGB was 0.2 for oil palms only (Alometri, KOROM, PHUA, & MATSUURA, 2016), which is almost twice less than what we found for date palm in the UAE. This could be attributed to the fact that desert ecosystems support accumulating less moisture in their plants.

Age is a determinant factor in accumulating biomass in date palm. In young palms, where there is no developed trunk, the average AGB is equal to 22.51 kg/palm. Then AGB is increasing dramatically in medium age palm by the effect of the trunk development and the increase in crown fronds numbers diameter to reach the average weight of AGB to 94.69 kg/palm. The increase in AGB is continuing in mature palms to average around 284.41 kg/palm. The percentages of AGB to total biomass is varying from 75.11% to 51.94% for oil palms respectively with an average of 59.34% for medium and mature palms alone because of the already developed trunk.

The ratio of BGB to AGB in date palm of Abu Dhabi is 0.332 in young date palms; it continues to increase to reach 0.925 in medium date palm then decreases to stabilize at around 0.496 for mature palm. This increase of BGB for medium date palm although paralleled by the emergence of the trunk however is offset by the rapid increase in the root system. Still the ratios in all age classes of date palm are very high and exceeding largely the 0.2, which is the general ratio recommended by many researchers (Achard et al., 2002; Gibbs et al., 2007; Houghton, Hall, & Goetz, 2009). However, we have to consider that these researchers have derived there ratios from regular trees not palms and also from tropical, boreal and temperate ecosystems, not from desert vegetation where there is no known or published BGB to AGB ratios (Mokany et al., 2006).

Table 5: Mean, minimum and maximum DNs of 3X3 pixel window for date palm plots, (a) Raw bands and (b) VI's,

<table>
<thead>
<tr>
<th></th>
<th>B1</th>
<th>B</th>
<th>G</th>
<th>R</th>
<th>NIR</th>
<th>SWIR1</th>
<th>SWIR2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min.</td>
<td>11776.6</td>
<td>11082.9</td>
<td>11078.4</td>
<td>11224.4</td>
<td>17685.3</td>
<td>14523.4</td>
<td>11254.7</td>
</tr>
<tr>
<td>Max.</td>
<td>13316.7</td>
<td>13140</td>
<td>14152.1</td>
<td>16317.6</td>
<td>20515.7</td>
<td>22286.2</td>
<td>19159.2</td>
</tr>
<tr>
<td>Mean</td>
<td>12351.9</td>
<td>11873.4</td>
<td>12260.1</td>
<td>13204.7</td>
<td>19221.3</td>
<td>17709.6</td>
<td>14606.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>DVI</th>
<th>GNDVI</th>
<th>NDI</th>
<th>NDVI</th>
<th>RVI</th>
<th>SAVI</th>
<th>SR</th>
<th>TCB</th>
<th>TCG</th>
<th>TCW</th>
<th>TVI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min.</td>
<td>3972.22</td>
<td>0.18</td>
<td>-8.49</td>
<td>0.11</td>
<td>0.56</td>
<td>0.17</td>
<td>1.25</td>
<td>32076.4</td>
<td>-2919.49</td>
<td>-7496.36</td>
<td>1.05</td>
</tr>
<tr>
<td>Max.</td>
<td>8623.56</td>
<td>0.28</td>
<td>-0.17</td>
<td>0.28</td>
<td>0.8</td>
<td>0.42</td>
<td>1.78</td>
<td>42035.9</td>
<td>1595.99</td>
<td>-1243.99</td>
<td>1.13</td>
</tr>
<tr>
<td>Mean</td>
<td>6020.89</td>
<td>0.22</td>
<td>-3.71</td>
<td>0.19</td>
<td>0.69</td>
<td>0.28</td>
<td>1.47</td>
<td>35757.6</td>
<td>-762.008</td>
<td>-4188.48</td>
<td>1.09</td>
</tr>
</tbody>
</table>

Where B1 is band1, B is band2 band, G is band3, R is band4, NIR is band 5, SWIR1 is band6, and SWIR2 is band7.
The high correlation between age and biomass (around 0.95 with P-value < 0.01), makes age as the best indicator to estimate AGB of date palm. Results showed that date palm has three age classes regarding biomass storing: 1) Young age palm where most of the biomass is stored in the crown of the date palm with about 75.11% of the total biomass; 2) Medium age palm where trunk starts developing in date palm with an increasing portion from the total biomass approaching 16.2%. On the other hand, the root biomass ratio is doubled from 24.89% to 48.06% from the total biomass and; 3) Mature age palm where ratios of the three parts of date palm approach equilibrium with a value of one third for each. The regression analysis of age with AGB of the date palm is showing that crown biomass increasing rate is around 14% yearly and trunk biomass increasing rate is around 18% yearly.

Age is an unstructured variable and cannot be measured directly on the field without asking the farmers to know the history of the farm. Therefore, using structural variables, (e.g height, trunk height and DBH) to build AGB regression equations that can be further correlated with remote sensing variables is highly suitable (R. Hereward V. Corley & Tinker, 2008a; Khalid et al., 1999; ). Generally, DBH is widely used in biomass equations in tropical regions because of the high correlation between DBH and AGB (S. Brown, 1997). For palms, it is recognized that the correlation between DBH and biomass is weak to intermediate. We found in our study that the correlation between DBH and crown biomass is 0.33 and not significant (> 0.05) and the correlation between age and trunk biomass is intermediate with 0.6. This due to the growing nature of the palm trunk where the DBH becomes almost stable and there is no significant increase in DBH from medium to mature date palms.

Regarding both parts of AGB (crown biomass and trunk biomass) of date palm, the best correlation was found to be height, trunk height, frond numbers, crown area and crown diameter of date palms (< 0.01) (age was excluded because this is a nonstructural variable). The coefficient of determination (R square) was used to find the best predictors for both crown and trunk biomass; we concluded that the best values of R square were found to be: (R² = 0.8838) for crown biomass prediction (table 3); and (R² = 0.8354) for trunk biomass prediction (table 4).

5. CONCLUSION & RECOMMENDATIONS

The aim of this paper was to build up allometric equations of date palm as species. However, the researchers concluded that there is a role of date palm cultivars in accumulating biomass and carbon sequestration. In oil palms, there is considerable variation between palms even within the same cultivar and there are large and statistically significant differences between progenies for all characters measured. Therefore, it is so important to study this subject in future research to have full idea about the best date palm cultivars for carbon sequestration.

- For date palm of Abu Dhabi, structural field variables such as: crown area (CA) and trunk height (Ht), can be used for estimation of AGB (= crown biomass plus trunk biomass).
- Crown biomass can be estimated using crown area (in m²) as a predictor by the following exponential regression equation:
  \[ \text{Crown Biomass} = 14.034 \times 1.057^{\text{CA}} \ (with \ R^2 = 0.8838) \]  
  (7)
- Trunk biomass can be estimated using trunk height (in m) as a predictor by the following polynomial regression equation:
  \[ \text{Trunk Biomass} = -3.956(\text{Ht})^2 + 55.247 (\text{Ht}) - 2.0342 \ (with \ R^2 = 0.8354) \]  
  (8)
- BGB can be estimated as a ratio of AGB by using different ratios according to the palm age (0.332, 0.925 and 0.496 for young, medium and mature date palms respectively).
- The study found that the ratio of BGB to AGB in date palm in the desert ecosystem is much higher than what is existing in previous studies with other species and ecosystems. This issue must be investigated more as the researchers were aware of the errors that could be involved as some of the root systems considered in the analysis involved the accumulation of certain amounts of soils.

ACKNOWLEDGMENT

The researchers would like to express their appreciation to the Research Affairs - United Arab Emirates University for their financial support under fund # 31S247. The researchers would also like to express their gratitude to Dr.
Mohamed T. Moussa for his ecology Lab assistance and to Al Foah farm company for providing us free access to their premise and to allow us to use their big scale balance as well as their logistic facilities.

REFERENCE


USE OF ARTIFICIAL NEURAL NETWORKS FOR ESTIMATING WINTER WHEAT LEAF AREA INDEX WITH SENTINEL-2/MSI IMAGERY

Yuanheng Sun (1), Qiming Qin (1), Huazhong Ren (1), Zhaoxu Zhang (1), Zehao Long (1)

1 Institute of Remote Sensing and GIS, School of Earth and Space Science, Peking University, No. 5, Yiheyuan Road, 100871 Beijing, China
Email: yhsun@pku.edu.cn; qmqin@pku.edu.cn; renhuazhong@pku.edu.cn; 1601110526@pku.edu.cn; longguoxxwl@163.com

KEY WORDS: leaf area index, Sentinel-2, winter wheat, neural network, solar zenith angle

ABSTRACT
As a new generation of satellite launched by ESA, Sentinel-2 could achieve a high-frequency ground observation in a 5-day revisit period with A/B two satellites being in orbit. The satellite-equipped MSI (Multi-Spectral Instrument) sensor provides multi-spectral imagery with spatial resolutions of up to 10 m, which marks a new source of data for dynamic monitoring of crop growth. In this paper, an artificial neural network (ANN) algorithm for winter wheat leaf area index (LAI) inversion is proposed. First, the winter wheat canopy spectra were generated based on the PROSAIL model in multiple sets of single solar zenith angles (SZA). Then, the corresponding broad-band reflectance is obtained by integrating with the MSI spectral response as the ANN training data. For 10-m resolution LAI retrieval, the surface reflectance of blue (B2), green (B3), red (B4) and near-infrared (B8) bands are used as the inputs of the ANN training. In the LAI inversion process, two nearest SZA values in network collection will be searched out for each pixel first, and the LAI will be estimated by these two networks, separately. The final LAI estimation result will be the linear interpolation of the retrieval outcomes based on their SZA value. The performance of our proposed model is evaluated and validated by using the simulated data generated by the PROSAIL model and the ground measured data afterwards. The evaluation based on simulated data demonstrated that the ANN model proposed by this paper performs better than the traditional ANN inversion method, with RMSE decreasing from 0.37 to 0.20. Validation against ground LAI observation on Sentinel-2 imageries achieves an acceptable accuracy, with \( R^2 = 0.90 \) and RMSE = 0.73, which will benefit the agriculture applications. Future study for LAI estimation will focus on the use of extra red edge bands of Sentinel-2 imagery in order to achieve a higher accuracy.

1 INTRODUCTION
As the description of the canopy structure of crops in agroecosystem, leaf area index (LAI) is a key parameter of plant productivity and functioning, which also influence the photosynthesis process during the entire growing seasons (Asrar et al., 1984). Remote sensing offers a timely and convenience approach to obtain the LAI of crops, and this gradually proves to be the best option for LAI acquisition in vast areas. For precision agriculture, the demand of spatial resolution is at least 20 m, or perhaps 10 m better for site-specific management (Mulla, 2013). As a result, the high spatial resolution remotely sensed imagery is of urgent need for agricultural application. Meanwhile, the timely acquisition of crop growth parameters in the growing season is also of great importance for the application in precision agriculture. As a new generation satellite launched by European Space Agency (ESA) in 2015 and 2017, Sentinel-2 could achieve a high-frequency ground observation in a 5-day revisit period with A/B two satellites being in orbit (Drusch et al., 2012). The satellite-equipped MSI (Multi-Spectral Instrument) sensor provides multi-spectral imagery with a spatial resolution of up to 10 m, which make it an ideal data source for crop LAI dynamic monitoring in precision agriculture.
Using multispectral earth-observation imagery, the method of LAI retrieval could mainly grouped as statistical based model, radioactive transfer based model and data driven based model. Statistical method is simple and easy to operate, but it require site and biome specific calibration, which limit its usage in large region (Carlson et al., 1997). Radioactive transfer based model simulate the movement of photons within a canopy, thus is relative accurate (Myneni et al., 2002; Verget et al., 2011). However, it is usually very complicated, time consuming and not easy to be applied in operational LAI production. To resolve this issue, Data driven model attracts much attention in recent year due to the development of machine learning algorithm and remote sensing imagery accumulation. However, huge amount of LAI observation data is required for model training and validating, which restricts its further development. As a result, many studies utilize the radioactive transfer model to simulate vegetation canopy spectra, which are regarded as training data for machine learning model (Duan et al., 2014; Li et al., 2018).

This work aim at the estimation of winter wheat LAI with Sentinel-2/MSI Imagery on the basis of artificial neural networks. The neutral networks are grouped by solar zenith angle (SZA) and are supposed to construct with the simulated data generated by PROSAIL model on parameters of winter wheat and Sentinel-2/MSI spectral response. Finally, the network performance is evaluated and validated on simulated data and in situ LAI observations.

2 MATERIAL AND METHODOLOGY

2.1 Simulated spectra

The simulated dataset was generated on the basis of PROSAIL radiative transfer models in leaf and canopy scales, respectively (Berger et al., 2018). The input parameters and their nominal value or ranges of the model were shown in Table 1 according to related literatures (Li et al., 2018) and our field measurements.

<table>
<thead>
<tr>
<th>Input parameter</th>
<th>Nominal value or range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solar zenith angle</td>
<td>0-60 (step by 5)</td>
</tr>
<tr>
<td>Observer zenith angle</td>
<td>0</td>
</tr>
<tr>
<td>Leaf area index</td>
<td>0-7</td>
</tr>
<tr>
<td>Average leaf angle</td>
<td>57</td>
</tr>
<tr>
<td>Hot spot</td>
<td>0.1</td>
</tr>
<tr>
<td>chlorophyll content</td>
<td>35-65</td>
</tr>
<tr>
<td>Water content</td>
<td>0.01-0.03</td>
</tr>
<tr>
<td>N (Leaf mesophyll structure index)</td>
<td>1.5</td>
</tr>
<tr>
<td>Dry matter content</td>
<td>0.005-0.011</td>
</tr>
</tbody>
</table>

Besides, the background soil was randomly selected from Johns Hopkins University (JHU) Spectrum Library. For each solar zenith angle value, a total of 10,000 cases were simulated with a random combination of other variables. To account for the instrumental and atmospheric noise, multiplicative uncertainties were added based on white Gaussian noise to the simulated spectra:

\[ R_{mol}(\lambda) = R_{sim}(\lambda) \left[ 1 + g(0, \sigma(\lambda)) + g(0, \sigma(all)) \right] \]

where \( g(0, \sigma) \) represents a normal distribution (mean = 0 and variance =\( \sigma^2 \)) and \( \sigma(\lambda) = \sigma(all) = 4\% \). Afterwards, broad-band surface reflectance of Sentinel-2 was obtained by the simulated spectra integrating with the spectral response function of Sentinel-2 MSI.

2.2 Neural network construction

Back-propagation (BP) neural network is often adopted to deal with the relevant works, it includes an input layer, hidden layer, and output layer, as well as network initialization (i.e., the number of neurons is determined by the input and expected output to initialize weights between neurons), hidden layer, and output layer calculations. The
error values and weights are updated to obtain the final weight.

The simulated surface reflectance of blue, green, red and near infrared bands with a spatial resolution of 10 m were selected as the input variables of the BP neural network. The inputs and output are normalized to prevent possible numerical problems in the training process. Normalization is achieved by scaling between -1 and +1 according to the following expression.

\[
x_{\text{norm}} = 2 \frac{x - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}} - 1
\]

where \( x \) represents the inputs or output, \( x_{\text{min}} \) and \( x_{\text{max}} \) are the minimum and maximum values of \( x \) respectively. Once the minimum and maximum values of each variable are calculated, they are fixed afterwards. Finally, \( x_{\text{norm}} \) is the corresponding normalized result. One hidden layer with 5 tangent-sigmoidal neurons are set to obtain an optimal balance of training accuracy and time consuming. The overall structure of BP neural network is shown blow.

```
Figure 1. Structure of BP neural network
```

Network for each SZA group is trained independently. In the predicting process, 2 networks of SZA group are selected based on input SZA and nearest neighbor principle. The final estimated LAI result will be the linear interpolation of 2 reversed LAI values with the selected networks.

### 3 RESULTS AND DISCUSSION

#### 3.1 Evaluation with simulated data

The theoretical performances of method proposed by this work were evaluated over a new test simulated dataset. The SZA of the evaluation dataset follows a uniform distribution from 0 to 60, other than single value in training one. For comparison, a single BP neural network was constructed by another dataset whose SZA varied from 0 to 60 and other parameters stayed in constant with evaluated dataset.

It shows from Figure 2 that both method were efficient for LAI estimation. The dispersion of our method around 1:1 line is very small over the whole range of variation of LAI. Moreover, it achieves a smaller RMSE of 0.20 compared with 0.37 of single ANN method. However, it is noted that a few underestimations are observed in the scatterplot of our model. This is because failed estimation is assigned 0, thus it will bring down the overall result after the weighted average process.
3.2 Validation with Sentinel-2 data

The field measurement of winter wheat LAI was conducted on Mar 28th and May 4th in Hengshui, Hebei Province of China, which nearly in accordance with the time of Sentinel-2 imagery acquisition. Five LAIs were observed for an elementary sample unit (ESU) and the mean value was regarded as its final result. The imageries were downloaded from the Sentinels Scientific Data Hub (http://scihub.copernicus.eu/) as Level-1C TOA reflectance. Atmosphere correction was conducted using the Sen2Cor atmosphere correction toolbox (version 2.4.0) inbuilt in the Sentinel Application Platform (SNAP) software (version 5.0.0) to obtain the surface reflectance.

Figure 3 shows part of LAI estimation result in our study area. The LAI is approximately 1-2 in late March and varies 3-5 in late April, on the basis of our estimating outcomes. A good agreement between the LAI estimates obtained from our ANN retrieval method and from field observations is observed for most ESUs, with an overall $R^2=0.90$ and RMSE=0.73. The underestimate is observed for most dots in Figure 4, but more significant in large LAIs. There are 6 days delay for field measurement compared with Sentinel-2 imagery, thus lead to this underestimated result especially in crop growing seasons.
Figure 4. Comparison of estimated and measured LAI based on the ANN method proposed.

4 CONCLUSION

This work proposed an ANN method of winter wheat LAI retrieval for Sentinel-2 MSI imagery. The method was evaluated and validated by simulated data and in-situ observations separately, which achieved acceptable accuracy. Future study will focus on the use of extra red edge bands of Sentinel-2 imagery in order to achieve a higher accuracy.

REFERENCES


REMOTELY SENSED RELATIVE HUMIDITY FOR PREDICTING METISA PLANAS’S POPULATION OIL PALM PLANTATIONS

Siti Aisyah Ruslan (1), Farrah Melissa Muhamram (1), Dzolkhifli Omar (2), Zed Diyana Zulkafli (3), Muhammad Pilus Zambri (4)

1Department of Agriculture Technology, Faculty of Agriculture, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor Darul Ehsan, Malaysia
2Department of Plant Protection, Faculty of Agriculture, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor Darul Ehsan, Malaysia
3Department of Civil Engineering, Faculty of Engineering, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor Darul Ehsan, Malaysia
4Department of Agronomy and Innovation, TH Plantations Berhad, Level 31 – 35, Menara TH Platinum, No. 9, Jalan Persiaran KLCC, 50088 Kuala Lumpur, Malaysia

Email: aisyahruslan01@gmail.com; farrahm@upm.edu.my; dzolkhifli@gmail.com; zeddiyana@upm.edu.my; pilus@thplantations.com

KEY WORDS: Metisa plana; Relative humidity; Outbreak prediction; Artificial neural network; Geospatial technology

ABSTRACT: Metisa plana (Walker) is leaves defoliating insect that is able to cause a staggering loss of USD 2.32 billion within two years to Malaysian oil palm industry. Therefore, an early warning system to predict the outbreak of Metisa plana that is cost, time, and energy effective is crucial. In order to do this, the role of environmental factors such as relative humidity (RH) on the pests’ population’s fluctuations should be well understood. Hence, this study utilized the geospatial technologies to i) to construct the relationship between the geospatially derived relative humidity and Metisa plana outbreak, and ii) to predict the outbreak of Metisa plana in oil palm plantation. Metisa plana census data of larvae instar 1, 2, 3, and 4 were collected approximately biweekly over the period of 2014 and 2015. Moderate Resolution Imaging Spectroradiometer (MODIS) satellite images providing values of RH were extracted and apportioned to 6 time lags; 1 week (T1), 2 weeks (T2), 3 week (T3), 4 weeks (T4), 5 week (T5) and 6 weeks (T6) prior to census date. Pearson’s correlation, multiple linear regression (MLR) and multiple polynomial regression analysis (MPR) were carried out to analyse the linear relationship between Metisa plana and RH. Artificial neural network (ANN) was then used to develop the best prediction model of Metisa plana’s outbreak. Results show that there are correlations between the presence of Metisa plana with RH, however, the time lag effect was not prominent. MPR was able to produce model with higher R² in comparison to MLR with the highest R² for both analysis were 0.48 and 0.15 respectively at T4 to T6. Model with the highest accuracy was achieved by ANN that utilized the RH at T1 to T3 at 95.29%. Based on the result of this study, the prediction of Metisa plana’s landscape ecology was possible with the utilization of geospatial technology and RH as the predictor parameter.

1. INTRODUCTION

Untreated infestation of Metisa plana can lead to a devastating losses in oil palm industry mainly due to the potential of causing a complete skeletonization and an eventual death of oil palm fronds. This will jeopardize the ability of palms to carry out photosynthetic activities that will suppress the growth of the palms and reduce the yield. Wood et al. (1973) earlier captured the threatening nature of this pest by demonstrating that 50% of damage caused by its infestation will bring about 43% or approximately 10 t ha⁻¹ of fresh fruit bunch (FFB) for the next two consecutive years. This was later supported by Basri (1993) which stated that even at a lower level of damage of 10% to 13%, the loss of yield of oil palm can go up to 40%. These figures can be translated into a loss of approximately USD 2,032 per hectare, which will eventually cause a loss of USD 2.32 billion for two consecutive years, given only 10% of the 5 million hectares of oil palms in Malaysia being infested.

Taking into account the high potential of severe economic losses that could be initiated by Metisa plana, rigorous control methods and mitigation systems should be properly planned and executed. Furthermore, the control methods must be efficient in terms of cost, time, and man power, as well as not harmful to oil palms and the environment. The success in
Relative humidity is the derivation of temperature and moisture from rainfall which significantly affects insect population dynamics. The influence of relative humidity on insect development and survivability is especially important during early developmental stages of insects. Higher level of relative humidity has been found to be beneficial for normal embryonic development as well as to promote faster hatchability for some insects’ eggs (Godfrey et al., 1991; Ofomata, et al., 2000; Tamiru et al., 2012). However, on both extreme ends, relative humidity may adversely affect eggs hatchability. For instance, an insufficient level of humidity is able to hinder the release of larvae from eggshell due to the loss of lubrication and hardened cuticle caused by desiccation (Guarneri et al., 2002). As such, excessive humidity, on the other hand, was determined as one of the cause for egg mortality due to drowning and pathogen infection (Guarneri et al., 2002; Gullan and Cranston, 2004; Hirose et al., 2006; Norhisham et al., 2013). Relative humidity (RH) has also been associated with entomopathogenic fungal infections among insects. The abundance of fungal infections among insects and their larvae are greater, parallel to the increase of relative humidity (Luz and Fargues, 1997; Shipp et al., 2003; Mishra et al., 2013). In a study conducted by (Sajap and Siburat, 1992), high level of relative humidity was positively correlated with the infections of fungus in bagworm, where, a higher number of infected bagworms were found to reside on the middle and the lower tree canopies where the relative humidity was higher in comparison to the upper level of tree canopies.

Currently, the effort of understanding ecological aspects affecting the insect pest outbreaks, along with their control practices are still leaning towards the conventional approaches that are highly dependent on in-situ data collection which can be ineffective. The exploitation of modern technology such as geospatial technology have essentially benefited the agricultural industry. This technology, which encompasses of remote sensing, global positioning system (GPS), and geographic information system (GIS) have been used in the assessment and monitoring of crop pests and diseases. Unlike the conventional approaches, information on the triggering factors of insect pest outbreaks such as temperature, rainfall, and vegetation’s condition can be obtained through this technology in rapid, harmless, and cost-effective manners.

ANN is a sophisticated computer program in which its working principle imitates the way human brain process information i.e. through pattern recognition and relationship determination. Unlike the other computer programs that process information through programming, ANN conduct information processing and knowledge gathering through experience i.e. learning and training of data. Due to its sophisticated way of computing and processing information, ANN has the ability to produce prediction model with higher accuracy compared to regression analysis (Sahin et al., 2013; Lee et al., 2016; Mishra et al., 2017) as its advanced mechanism allows it to excel in capturing non-linear relationships, which is limited for regression analysis to conduct.

Enlightened by the potential devastating impacts of Metisa plana towards the oil palm industry in Malaysia and the importance of the application of technologies in controlling them, the objective of this paper is to study the relationship between the presence of Metisa plana and geospatially derived relative humidity, and to predict the presence of Metisa plana using relative humidity as the predictor parameter using ANN.

2. MATERIALS AND METHOD

2.1 Study area

This study was conducted in Tabung Haji plantation in Sungai Mengah, Muadzam Shah, located at 2° 57’ 30” N, 102° 53’ 0” E to 3° 1’ 0” N, 102° 53’ 0” E in the state of Pahang, Malaysia (Figure 1). The study site is spread over an area of 2000 ha and is divided into two main divisions which are division A and B. The former division comprises of 10 blocks and the latter division contains 16 blocks. Stemming from the tropical climate of the region, the temperatures of this study area are fluctuating from 24°C to 35°C with the average rainfall of 1900 mm to 2500 mm annually and average relative humidity of 59% to 65%. The main cultivation of this area is oil palm crops ranging from 10 to 20 years old with bagworm’s infestation level ranging from zero to mild. This plantation incorporated IPM in its management practices with pesticides being applied through trunk injection when the bagworm infestation is above the economic threshold level.
2.2 Data collection

Bagworm census data

The census data were collected over the period of 2014 and 2015, with 21 and 25 census cycles were done in 2014 and 2015, of which each census date was approximately 2 weeks gap between each other. For each cycle, 25 palms were chosen randomly. The census was done in a destructive manner through the cutting of the frond number 17 for each of the palm chosen as it is considered to represent oil palm crown as a whole (Tailliez et al., 1992). Only bagworm larvae of instar 1, 2, 3, and 4 were collected due to the fact that bagworm larvae at these stages bring the most destruction to oil palm owing to the high level of leaf consumption. Furthermore, the vulnerability of bagworm larvae of instar 5 and above will be decreasing because of the reduction of leaf consumption paired with the harden of their outer cases making them less destructible by the insecticides (Kok et al., 2011). Recorded bagworm’s larvae for each palm were pooled and then averaged for each block.

Geospatial data

Since there was no direct retrieval method of RH from satellite measurements, the computation of RH was done according to algorithm described and modified for Malaysia’s utilization by Peng et al. 2006. Hence, in doing this, two sets of MODIS products were downloaded from http://ladsweb.nascom.nasa.gov/, a website maintained by The Level-1 Atmospheric Archive & Distribution System (LAADS) Distributed Active Archive Centre (DAAC). The first product was MOD07, a product that was responsible for the acquisition of surface air temperature (Ta) data used in Equation 1c, while the second product was MOD021KM that was responsible for the acquisition of 5 NIR bands used to compute precipitable water vapour (PW). Both of their temporal resolution were 1 day with spatial resolution of the first product being 5 km and second product being 1 km. Additional data used in RH computation was digital elevation model (DEM) over Malaysia from SRTM that was downloaded from https://earthexplorer.usgs.gov/.

A total of 444 images of each datasets for year 2014 and 443 images for year 2015 were downloaded. Firstly, the required band layers were extracted from each images by using MODIS conversion toolkit, an extension in ENVI software version 5.2 (Exelis Visual Information Solutions Inc., Boulder, Colorado, USA). Surface air temperature layers were extracted from MOD07 product images, while layers of NIR bands were extracted from MOD021KM product images. All of the layer spatial projections were set to WGS 1984 UTM Zone 47 N and the algorithm was applied to the layers to compute RH images. Then, all of the images were subset according to the study area.
2.3 Data extraction

Prior to data extraction, the RH layers pixel size were set to 250 m through resampling tool in ArcGIS software version 10.3.1 (Esri Inc., USA). Area-weighted mean for all of the images were then calculated for each block of the study area according to Equation 1. Later, the extracted data were apportioned to datasets according to census cycle. Each datasets consist of a pairing of the average number of bagworm of each block for each census cycle with the area-weighted mean of RH at 6 different time lags; 1 week (T1), 2 weeks (T2), 3 week (T3), 4 weeks (T4), 5 week (T5) and 6 weeks (T6) prior to census date.

\[
\text{Areal weighted mean} = \frac{\sum (\text{pixel area within a block} \times \text{pixel values within a block})}{\text{total area of each block}}
\]  

(Equation 1)

2.4 Data analysis

Statistical analysis

Statistical analysis was conducted by using statistical analysis software SAS version 9.2 (SAS Institute Inc., Cary, North Carolina, USA). The correlation between average bagworm per block with area-weighted mean of RH of each block of time lag T1, T2, T3, T4, T5, and T6 was determined by Pearson’s correlation coefficient (r) analysis using PROC CORR. The analysis was later followed by regression analysis (R²) in order to measure the magnitude of influence that independent variables possessed over the dependent variable. This analysis was conducted by using PROC REG in SAS software. Two types of regression analysis was conducted in this study. They were multiple linear regression analysis (MLR) and multiple polynomial regression analysis (MPR). RH data were apportioned into datasets of week 1 to 3 (T1 to T3) and week 4 to 6 (T4 to T6) and were used in this analysis.

**Artificial neural network (ANN)**

The structure of neural network analysis comprises three main layers that are the input, output, and hidden layer. It is worth noting that the input layer represents the independent variables, the output layer represents the dependent variables, and the hidden layer represents the relationship developed between input and output variables. The establishment of the relationship between the input and the output variables is done through a network of iterative training where each of the input carries a weight into the network. Each training set will then generate estimated output values that will be compared to the actual output values producing error. This error will be used in network training and the iterative process will be conducted until the minimum error is achieved (Ahmadi et al., 2017).

ANN was used to further analyse the relationship between dependent and independent parameters to further construct the prediction model for the outbreak of Metisa plana in this study. For this, the Alyuda NeuroIntelligence software version 2.2 (2001-2005 Alyuda Research, Inc.) was used. RH data were apportioned into datasets of week 1 to 3 (T1 to T3) and week 4 to 6 (T4 to T6), prior to the census date and these datasets was used to predict the number of bagworm occurrence. Datasets containing missing values was excluded before these data were exported to the Alyuda software. Network trainings were performed by using 7 different training algorithms; i) Quick propagation (QP), ii) conjugate gradient descent (CDG), iii) quasi- newton (QN), iv) limited memory quasi-newton (LMQN), v) levenberg-marquardt (LM), vi) online back propagation (OBP), and vii) batch back propagation (BBP). However, only training algorithm that produced the best results was presented in this study. The network architecture was optimized for more than 10,000 times in order to obtain the best training, testing and testing accuracy.

3. RESULTS AND DISCUSSION

Pearson’s correlation analysis conducted showed an absence of a linear relationship between the presence of Metisa plana and RH suggesting that the relationship is non-linear. Figure 2 had demonstrated an inconsistent pattern of the positive and negative correlations between RH and Metisa plana. The presence of Metisa plana were found to be positively and negatively correlated with both relatively low and high level of RH which shows that correlation analysis was not able to determine the optimum values of RH that favour Metisa plana’s population.
These results would indicate that through correlation analysis, the effect of different time-lag of RH towards the presence of *Metisa plana* is not pronounced, as the correlation occurrences between RH of different time lags and the presence of *Metisa plana* are more or less the same i.e. no one distinct time lag was observed to have a significantly more correlations than the other.

**Metisa plana**’s prediction model were generated using MLR, MPR, and ANN. An overall result demonstrated that the highest $R^2$ values were obtained through analysis done by ANN, followed by MPR, while MLR produced models with the lowest $R^2$ (Table 1). MLR analysis essentially predicts the relationship between dependent ($y$) and independent ($X_1, X_2...X_k$) variables by assuming that their associations are linear (Equation 2). MPR, on the other hand, were used to fit a non-linear data into least squares regression model that still assumes the linearity of the relationships when predicting the associations of dependent and independent variables (Equation 3).

$$y = a + \beta_1X_1 + \beta_2X_2 ... \beta_kX_k$$  \hspace{1cm} (Equation 2)

$$y = a + \beta_1X + \beta_2X^2 + \beta_3X^3 ... \beta_kX^k$$  \hspace{1cm} (Equation 3)

Whereas $y$ represents dependent variables i.e. bagworm numbers and $X$ represents independent variables i.e. LST, RF, and RH, $a$ represents initial intercepts and $\beta$ represents partial regression coefficients of variable $X$.

As a result, this principle limits the ability of MLR and MPR to predict non-linear relationships efficiently. ANN, alternatively, possess the capability to capture non-linear relationship and thus, is more efficient in predicting the relationship between dependent and independent variables in this study. This is in alignment with several studies that have demonstrated the ability of ANN to predict the relationship between dependent and independent variables more accurately when compared to MLR (Sahin et al., 2013; Lee et al., 2016; Mishra et al., 2017).

**Figure 2:** Correlation analysis between relative humidity and number of *Metisa plana*
Table 1: Comparison of $R^2$ values obtained from regression analysis and artificial neural network

<table>
<thead>
<tr>
<th>Predictor variables</th>
<th>MLR</th>
<th>MPR</th>
<th>ANN</th>
</tr>
</thead>
<tbody>
<tr>
<td>RH (T1 to T3)</td>
<td>0.08</td>
<td>0.42</td>
<td>0.57</td>
</tr>
<tr>
<td>RH(T4 to T6)</td>
<td>0.15</td>
<td>0.48</td>
<td>0.56</td>
</tr>
</tbody>
</table>

$R^2$ values for MLR and MPR, along with ANN accuracies obtained from the aforementioned analysis (Table 1) have confirmed the previous claim that stated that the relationship between RH and *Metisa plana* was non-linear. Subsequently, this finding could be accredited to the characteristic of RH that embed both temperature and moisture (water vapour) effects in a single parameter whereby it is defined by the amount of moisture held by the air, relative to its maximum holding capacity at a certain temperature (Wylie and Speight, 2012). Therefore, the fluctuation of RH is highly influenced by temperature which affects the moisture holding capacity of the air, and rainfall that affects the amount of moisture in the air. RH has also been associated in successfully assisting the breed, grow, and dispersion of certain insects. Since insects are highly prone to desiccation, having the right level of humidity is vital for their survivability. Nevertheless, high level RH has been associated with the increase of pathogens and fungi attack (Wylie and Speight, 2012).

This is further supported by Sajap and Siburat (1992) where 90% of fungal infected of *Pteroma pendula* was found on the bottom and middle of tree canopies where the RH was higher than the top canopies.

From the ANN model obtained, RH prediction model was best utilised to predict the presence of bagworm of 1 to 3 weeks in the future with 95.29% for T1 to T3 and 81.96% for T4 to T6, respectively (Table 2). Since this study focused on the prediction of *Metisa plana*’s larvae of instar level one to four, one to three weeks prior to these stages marks the reproductive stages of *Metisa plana*. This is in alignment with the importance of RH in regulating the efficiency of insects’ reproduction. The optimum level of RH for eggs hatchability is crucial due to the need of lubrication and soft cuticle tissues for eggs to be successfully hatched (Guarnieri *et al*., 2002). RH, on both extreme ends, could hinder this process through desiccation or drowning of insect eggs (Guarnieri *et al*., 2002; Gullan and Cranston, 2004; Hirose *et al*., 2006; Norhisham *et al*., 2013). However, as the relationship between the presence of bagworm and RH is non-linear, the optimum RH values for infestations to occur could not be determined.

Table 2: ANN analysis between RH and *Metisa plana*

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Network architecture</th>
<th>Training algorithm</th>
<th>Training absolute error</th>
<th>Validation absolute error</th>
<th>Training accuracy</th>
<th>Validation accuracy</th>
<th>Test accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>RH at T1-T3</td>
<td>3-8-1</td>
<td>Quick propagation</td>
<td>0.56</td>
<td>1.14</td>
<td>99.56 %</td>
<td>93.63 %</td>
<td>95.29 %</td>
</tr>
<tr>
<td>RH at T4-T6</td>
<td>3-6-1</td>
<td>Conjugate gradient descent</td>
<td>0.70</td>
<td>1.42</td>
<td>98.09 %</td>
<td>95.42 %</td>
<td>81.96 %</td>
</tr>
</tbody>
</table>

4. CONCLUSION

The best model obtained for a single variable predictor was RH of T1 to T3 with 95.29% accuracy. Paired with the spatial analysis of *Metisa plana*’s infestations’ distributions, these models would allow for an early warning system to be developed that would assist the decision making process by plantation’s managements in controlling *Metisa plana*’s outbreak. This system would provide them with a quantitative information on the potential magnitude of *Metisa plana*’s infestations in their plantations earlier than the conventional practices timeline at the specific potential location, and thus, permit them to strategies for the most efficient controlling method that would save cost, man power, and time.

On the whole, the application of remote sensing and geospatial technology with the integration of ANN prediction method were able to determine the factors influencing the outbreak of *Metisa plana*. Furthermore, these technologies were proven
to be reliable in predicting this pest’s prevalence through the construction of prediction models with the accuracies that go up to 95%.

ACKNOWLEDGEMENT

This research was funded by the Ministry of Higher Education Malaysia (MOHE). The authors would like to offer our deepest gratitude to Tabung Haji Plantation Berhad for the support in conducting this project.

This paper is derived in part from the article to be submitted to the JOURNAL OF COMPUTER AND ELECTRONICS IN AGRICULTURE, copyright Elsevier.

REFERENCES


Mapping Oil Palm expansion from 2000 to 2017 in Indonesia and Malaysia

Pegah Hashemvand Khiabani (1), Wataru Takeuchi (2)

1,2 Institute of Industrial Science, The University of Tokyo, Meguro 4-6-1, Tokyo 153-8505, Japan

Corresponding author: phkh1366@iis.u-tokyo.ac.jp

KEY WORDS: Remote sensing, Land cover map, pixel based analysis.

Abstract

Oil palm plantations are rapidly expanding in Malaysia and Indonesia, which leads to deforestation and other associated damages to biodiversity and ecosystem services. Therefore, in response to international criticism on oil palm practices, various certification schemes have been set up in recent years to monitor sustainable production of the oil palm but despite similar starting points and comparable rates of increasing productivity and profit in this sector, both countries have developed different policies and legislation regarding oil palm sustainable production. Respectively, it is important to understand and monitor how these counties are acting toward their commitments. Google Earth Engine (GEE) is a low-cost, accessible, and user-friendly tool for detecting the establishment and extension of industrial oil palm plantations which hosts publicly available satellite images and allows for land cover classification using inbuilt algorithms. This paper aims to use this tool to analyses trends of oil palm expansion at the national level in Indonesia and Malaysia as the result of different policies. In order to perform land cover classification different spectral bands (RGB, NIR, SWIR, TIR ) and optical indices were used from Landsat 5,7,8 image to distinguish the following land cover classes: mature oil palm plantation, immature oil palm plantation, forest, water, urban area, and Non Forest-Non Oil Palm. This study used ALOS PALSAR yearly mosaic to cloud cover limitation and spectral similarity of forest and oil palm plantation. Classification and Regression Trees (CART) algorithm has been used to classify land cover maps. In order to assess the accuracy of classification, this paper used overall accuracy and Kappa coefficient where the higher over accuracy at 68% was observed for using Landsat and PALSAR combination. Comparing the trend of the oil palm explanation from 2000 to 2017 in both countries, our results show that in both countries the plantation area has been drastically increased from 2000 to 2017, however the speeding of increasing has been reduced in both countries from 2010 because of assigning more policies into promoting oil palm intensification.

1.1 Background

Elaeis guineensis is one of the species of palms which is known as oil palm and is extensively cultivated in South-East Asia, especially in Indonesia, Malaysia, and Thailand, mainly because of its high oil-yielding capability in comparison to the other such as sunflower and rapeseed. As of 2016, the consumption of palm oil reached 35% of the global vegetable oil, and palm oil has become the most consumed vegetable oil (Khai et al. 2017). It is expected that the increasing global trend of palm oil consumption continues with a slower pace in the coming future years (Pirker et al. 2016). Nonetheless, without robust managed expansion and intensification, this industry can have significant negative environmental impacts, such as deforestation and loss of biodiversity (Cheng et al. 2018). According to Gunarso et al. (2013), oil palm expansion and land-use and land-cover change in Malaysia and Indonesia are the primary drivers of tropical deforestation and biodiversity loss. This industry also is one of the contributors of carbon emission through its impact on peatland ecosystems which are known as one of the great sources of carbon sinks in Peninsular Malaysia, Borneo, and Sumatra (Lee et al. 2016). In this regard, some certification schemes such as Roundtable on Sustainable Palm Oil (RSPO) (Von Geibler 2013), are aiming to evaluate the environmental, social and economic dimensions of oil palm cultivations and by these schemes, the palm oil should entirely be prepared in an environmentally and socially appropriate and economically beneficial way. RSPO conducts the certification scheme through a set of principles and criteria (Roundtable on Sustainable Palm Oil 2013) which are mainly targeting the environment and social dimensions of the oil palm cultivation.

Considering the global increasing trend of palm oil demand, on one hand, and international attempts on sustainable production of the palm oil, on the other hand, makes the absence of an accurate oil palm map especially in major producer countries undeniable.

Previous studies have discussed that Landsat TM remote-sensing images are difficult to acquire in the Tropics because of the frequent occurrence of thick cloud cover (Yu et al. 2013; Gong et al. 2016). This problem has led to the use of data obtained by Synthetic Aperture Radar (SAR), which is a sensor capable of all-day/all-weather observation for studies in the Tropics (Teng et al. 2015). Oil palms with
single trunk and large fronds give distinguishable backscatter properties in SAR data. Data from PALSAR (Phased Array type L-band Synthetic Aperture Radar), an instrument flown on the Japanese ALOS (Advanced Land Observing Satellite), which was launched on 24 January 2006 and was declared dead on 12 May 2011, have been used to map oil plantation areas (e.g. Morel et al. 2011; Morel, Fisher, and Malhi 2012; Koh, Miettinen, and Liew 2011; Gutierrez-Velez and DeFries 2013). Several studies integrated PALSAR and optical data to provide complementary features (backscatter properties and reflectance/emissivity characteristics) in oil palm mapping (e.g. Morel, Fisher, and Malhi 2012; Gutierrez-Velez and DeFries 2013; Li and Dong 2015). These methods of classifying oil palm land cover require training in remote sensing, expensive software to process satellite images, and expensive hardware with fast computer processing power and large storage capacities (Friess et al., 2011). While it is important that such mapping exercises be carried out cautiously. The advent of digital globes such as Google Earth has played an important role in facilitating public access to geospatial analysis and simple spatial analysis tools (Butler, 2006; Friess et al., 2011). Google Earth Engine (GEE) (http://earthengine.google.org) takes open source geospatial analysis one step further by providing a cloud computing platform for earth observation data analysis. It combines a public data catalogue, which consists of a nearly complete set of Landsat imagery from its start in 1972 until the present day, with a large-scale computational facility optimized for parallel processing of geospatial data (Hansen et al., 2013). GEE also hosts an imagery classification system in the cloud which enables one to run supervised learning algorithms across huge datasets in real time. These algorithms are trained to identify different land cover classes using hand-drawn points and polygons on the input dataset. This land cover classification method is rapid and accessible through the World Wide Web. Hence, GEE's computing infrastructure revolutionizes time-consuming remote sensing processes, facilitates access of remote sensing resources and tools to the public, and paves a new way forward for rapid land cover classification.

1.2 Objective
In this scope, this study defines its objective as mapping oil palm plantation in Malaysia and Indonesia with different complexity of landscapes in year between 2000 to 2017 using the low-cost, accessible and user-friendly Google Earth Engine's (GEE) cloud platform to compare the oil palm expansion in Malaysia and Indonesia, using Landsat and PALSAR data combination.

2. METHODOLOGY
2.1. Study area
Malaysia is located between 0° and 7° north and 99° to 109° east, with an area of 328,657 km². Malaysia climate is categorized as equatorial, being hot and humid throughout the year. The average rainfall is 250 centimeters a year and the average temperature is 27°C. Indonesia is located between 5° north and 10° south and 95° to 141° east, with an area of 1,811,569 km². Split by the equator, Indonesia has an almost entirely tropical climate, with the coastal plains averaging 28°C, the inland and mountain areas averaging 26°C, and the higher mountain regions, 23°C. The area's relative humidity is quite high, and ranges between 70 and 90 percent. Almost 70 percent of Indonesia's oil palm plantations are located on Sumatra where the industry was started during the Dutch colonial days. The remainder - around 30 percent - is largely found on the island of Kalimantan, Therefore, in this study in order to reduce computation time and process, Kalimantan and Sumatra from Indonesia and whole Malaysia considered as region of interest. Figure 1 shows the location of the region of the interest.

Figure 1. Region of interest, where green area shows territory of Indonesia and red shows territory of Malaysia. The red rectangle shows the area of interest of this study

2.2 Dataset
The dataset in the present study were 30-m resolution Landsat 8 surface reflectance Tier 1 yearly composite for years 2017 and 2015, 30-m resolution Landsat 7 and Landsat 5 surface reflectance Tier 1 yearly composite for years 2010, 2008, 2005, 2000 and 25-m resolution global yearly mosaic PALSAR/PALSAR2/JERS-1 dataset from years 2017, 2015, 2010 and 2008. In each year, for Landsat images, the images with the least cloud cover have been used to preform supervised classification. SAR dataset in this study, is in the Fine Beam Dual (FBD) polarization mode, i.e. dual polarization HH
(horizontal transmit and horizontal receive) and HV (horizontal transmit and vertical receive). The original HH, HV digital numbers were converted into backscattering coefficients (in decibels (dB)) considering equation 1.

$$\gamma_0 = 10 \log_{10}(DN^2) + k$$

where DN is the image pixel digital number measured in the SAR image (or more accurately, the average pixel value over a group of pixels). K is a calibration factor which varies depending on the SAR sensor and processor system used. For ALOS/PALSAR and ALOS-2/PALSAR-2 data provided by JAXA, the calibration factor is -83.0 db. Additional images for HH/HV and HH-HV were generated by respectively calculating the ratio of HH and HV and the difference of HH and HV. All SAR bands were filtered using Lee speckle Filter with 3 × 3 window size to reduce inherent speckle noise in PALSAR. The SAR dataset imagery was already ortho-rectificated and slope corrected using the 90 m SRTM Digital Elevation Model in GEE archive.

2.3 Mapping oil palm plantations

Google Earth Engine (GEE) (http://earthengine.google.org) by providing near almost completed Landsat archive and the other geospatial datasets and offering a large-scale cloud computing platform for earth observation data analysis, is significantly contributing in earth observation studies (Gorelick et al. 2017). GEE with its great computational potential for parallel processing enables earth scientists to run supervised learning algorithms on enormous datasets (Lee et al. 2016). This study used GEE’s potential in image classification to map oil palm plantations in Malaysia and Indonesia.

In order to compile all the selected could free images in to a single image, a median Reducer function was applied on the collection of the images. From the single Landsat image composite, this study calculated a set of four indices including the Normalized Difference Vegetation Index (NDVI; Equation (2)), the Soil-adjusted Total Vegetation Index (STVI; Equation (3)), the Normalized Difference Land Surface Water Index (NDWI; Equation (4)) and Soil Index (SI; Equation (5)) to provide better information of land surface for the classification. These indices are defined as:

$$NDVI = \frac{b_{NIR} - b_{RED}}{b_{NIR} + b_{RED}}$$  \hspace{1cm} (2)

$$STVI = \frac{b_{SWIR1} - b_{RED}}{b_{SWIR1} + b_{RED} + 0.1} * (1.1 - \frac{b_{SWIR2}}{2.0})$$  \hspace{1cm} (3)

$$NDWI = \frac{b_{GREEN} - b_{NIR}}{b_{GREEN} + b_{NIR}}$$  \hspace{1cm} (4)

$$SI = \frac{b_{RED}}{b_{SWIR1}}$$  \hspace{1cm} (5)

Where NIR is near infrared band (B5 in Landsat 8 and B4 in Landsat 7,5), RED is red band (B4 in Landsat 8 and B3 in Landsat 7,5), GREEN is green band (B3 in Landsat 8 and B2 in Landsat 7,5), and SWIR is shortwave infrared band (B6,5 in Landsat 8 and B5,7 in Landsat 7,5). These indices were selected regarding their demonstrated potential in forest and land cover mapping. Also LSWI (Land Surface Water Index), NDTI (Normalized Difference Tillage Index) and EVI (Enhanced Vegetation Index) have been shown to have great capability to predict different land cover and forests (Torbick et al. 2016), however, considering the complex surface characteristics in this study, these indices couldn’t contribute well in predicting the land cover type and therefore have been excluded from further processes.

This study plotted about 80 training points and 50 training polygons for each land cover classes of mature oil palm, immature oil palm, urban, bare soil, water body, forest, and other vegetation, and used 90% of the training data to train the classifiers over six bands of Landsat (three visible bands, one Near Infrared and two Shortwave infrared bands) and four optical indices while the remaining 10% were used to conduct accuracy assessments. CART was used as the main classifier in this study to perform land cover classification in a pixel-based approach. This classifier has shown a great capability in oil palm classification in a study done by Lee et al. (2016). To assess the performance of GEE’s classification, this study verified land cover map produced by GEE with a set of randomly selected training points from high resolution Google Earth images on Google Earth Pro platform.

3. RESULT AND DISCUSSION

3.1 Signature Analyses of Land Cover predictors
To obtain the desired classification result, it is necessary to understand the signatures of various land cover types. Therefore, to select the best predictors for different land cover types which provide better information of land surface for the classification, different indices from Landsat images and different polarization combination from PALSAR images have been tested. From Landsat images, NDVI, EVI, LSWI, NDWI, STVI, NDTI and SI have been calculated and then the reflectance signature of each classes analyzed for all indices. Figure 2 shows spectral signature of different land cover class regarding different indices where comparing NDWI and LSWI, water can be distinguished more clearly by taking NDWI. NDVI can perform better in compare to EVI to differentiate vegetation from the other classes. Bare soil can be clearly predicted using SI in compare to NDTI. Therefore, to train the classifier based on Landsat data, this study apart from six bands of Landsat (three visible bands, one Near Infrared and two Shortwave infrared bands) considers four indices of SI, NDWI, NDVI and STVI. Nevertheless, because of the similar reflectance of oil palm and forest, it is difficult to accurately differentiate oil palm from forest using Landsat (Figure 2). Having said that, optical remote sensing such as Landsat images are also vulnerable to the effects of clouds, which can limit data availability for the oil palm plantation areas in the humid tropics.

To overcome these limitations this study used PALSAR images which has shown robust performance in forest loss detection. To compare the difference between oil palm plantation and other land cover types, this study compare the signature of all land cover types considering HH, HV, HH/HV and HH-HV (Figure 3). Water has the lowest values in HH and HV because of absorption and mirror reflection, but the highest values in the difference and the ratio values, compared to the other five land cover types. Thus, water can be easily identified. Oil palm has HH and HV values between water and forest although closer to forest, but higher ratio values than those of forest. Note that the difference in the ratio values among different land cover types is of too small an order of magnitude to recognize. Therefore, it is not easy to pick up oil palm based on the HH/HV ratio values. Though oil palm and natural forest have partially overlapped HH data, oil palm has bigger HV backscatter data than natural forest, which is mainly caused by their differences in branch, trunk, and canopy. At the same time, natural forest has different difference values (HH-HV image) and is separable from oil palm plantations. Therefore, taking both the HV values and HH-HV difference values into account, oil palm plantations can be separated from natural forests and other type.

3.1. Oil palm plantation mapping

To identify oil palm plantation from year 2000 to 2017 in Malaysia and Indonesia, this study conducted classification on one year composite of Landsat 8 surface reflectance combined with PALSAR yearly mosaic for years 2017 and 2015, one year composite of Landsat 7 surface reflectance combined with PALSAR yearly mosaic for years 2010 and 2008 and one year composite of Landsat 7 and Landsat 5 surface reflectance for years 2005 and 2000. In this approach, GEE's CART classifier could detect oil palm plantations from Landsat 8. In particular, the CART classifier provided the overall accuracy of 68% and Kappa coefficient of 0.63 using combination of Landsat and PALSAR images. The overall accuracy of 50% and Kappa coefficient of 0.41 was calculated for classification using only Landsat images. The small and fragmented oil palm plantations which usually are cultivated by smallholders makes classification more challenging to identify the oil palm plantations, which result in relatively average kappa coefficient (Cheng et al. 2018). The producer’s accuracy for water body and forest have the highest value followed by urban area, mature oil palm and None forest None Oil palm, the lowest value assigned to immature oil palm and bare soil classes (Table 4) in the case of combination of Landsat and PALSAR. The extensive heterogeneous landscape also makes interpretation for sampling and classification more challenging which also effects the classification accuracy. Furthermore, fine resolution images in Google Earth Pro which have been used for accuracy assessment were often obtained from different years which considering land cover changes over the years, could be one of the sources of moderate classification accuracy.

The results of the mapping using the two different inputs (Landsat and PALSAR) for an area in Malaysia are shown in Figure 4 (using CART classifier). A visual inspection shows that the results of taking the Landsat alone and taking the Landsat and PALSAR combination classification are different. This observation is true for either Malaysia or Indonesia. Testing samples for both study areas were used to evaluate the detection accuracies. The testing data have been prepared from Google Earth Pro which visually interpreted. The overall accuracies of the Landsat and PALSAR combination were higher using CART classifiers than the results obtained using Landsat separately (see Tables 1 and Table 2). For a more detailed understanding of the precision of the interpretation of classification, a comparison was conducted using precision measurement indices based on a confusion matrix, including Producer’s Accuracy, and User’s Accuracy.
Figure 2 - Spectral Signature of different land cover class with different indices. a) Spectral Signature of different land cover class regarding LSWI. b) Spectral Signature of different land cover class regarding NDWI. c) Spectral Signature of different land cover class regarding EVI. d) Spectral Signature of different land cover class regarding NDVI. e) Spectral Signature of different land cover class regarding NDTI. f) Spectral Signature of different land cover class regarding SI. g) Spectral Signature of different land cover class regarding STVI.
As can be seen in Table 1 and Table 2, the Overall Accuracy of the Landsat and PALSAR combination using CART is the higher at 68% while those of Landsat separately are 50%. Based on the classification, the total area covered by oil palm plantation in Malaysia and Indonesia in years between 2017 to 2000 are summarized in Table 3.

Figure 3- Spectral Signature of different land cover class with different indices. a) Spectral Signature of different land cover class regarding HH. b) Spectral Signature of different land cover class regarding HV. c) Spectral Signature of different land cover class regarding HH/HV. d) Spectral Signature of different land cover class regarding HH-HV
Figure 4- Landsat 8 composite year 2017 and classification results. a) Landsat 8 surface reflectance yearly composite (2017). b) Classified image with CART classifier using Landsat. c) Classified image with CART classifier using Landsat and PALSAR.

Table 1- Landsat and PALSAR combination classification Confusion Matrix

<table>
<thead>
<tr>
<th></th>
<th>Mature Oil Palm</th>
<th>Immature Oil Palm</th>
<th>Bare Soil</th>
<th>Urban Area</th>
<th>Water</th>
<th>Forest</th>
<th>None-Forest None-Oil Palm</th>
<th>Total</th>
<th>Producer Accuracy</th>
<th>Omission</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mature Oil palm</td>
<td>13</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td>20</td>
<td>0.7</td>
<td>0.35</td>
</tr>
<tr>
<td>Immature Oil Palm</td>
<td>2</td>
<td>10</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td>20</td>
<td>0.5</td>
<td>0.50</td>
</tr>
<tr>
<td>Bare Soil</td>
<td>2</td>
<td>0</td>
<td>9</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>20</td>
<td>0.5</td>
<td>0.55</td>
</tr>
<tr>
<td>Urban Area</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>14</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>20</td>
<td>0.7</td>
<td>0.3</td>
</tr>
<tr>
<td>Water</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>17</td>
<td>2</td>
<td>0</td>
<td>20</td>
<td>0.9</td>
<td>0.15</td>
</tr>
<tr>
<td>Forest</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>18</td>
<td>1</td>
<td>20</td>
<td>0.9</td>
<td>0.1</td>
</tr>
<tr>
<td>None-Forest/None-Oil Palm</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>14</td>
<td>20</td>
<td>0.7</td>
<td>0.3</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>21</td>
<td>12</td>
<td>18</td>
<td>18</td>
<td>21</td>
<td>27</td>
<td>23</td>
<td>140</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

User Accuracy 0.62 0.83 0.50 0.78 0.81 0.67 0.61 Overall Accuracy=67.8%
Commission 0.38 0.17 0.50 0.22 0.19 0.33 0.39 Kappa Co.=0.63

Table 2- Landsat classification Confusion Matrix

<table>
<thead>
<tr>
<th></th>
<th>Mature Oil Palm</th>
<th>Immature Oil Palm</th>
<th>Bare Soil</th>
<th>Urban Area</th>
<th>Water</th>
<th>Forest</th>
<th>None-Forest None-Oil Palm</th>
<th>Total</th>
<th>Producer Accuracy</th>
<th>Omission</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mature Oil palm</td>
<td>6</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>3</td>
<td>20</td>
<td>0.3</td>
<td>0.7</td>
</tr>
<tr>
<td>Immature Oil Palm</td>
<td>6</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>5</td>
<td>20</td>
<td>0.1</td>
<td>0.9</td>
</tr>
<tr>
<td>Bare Soil</td>
<td>8</td>
<td>1</td>
<td>5</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>20</td>
<td>0.3</td>
<td>0.75</td>
</tr>
<tr>
<td>Urban Area</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>14</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>20</td>
<td>0.7</td>
<td>0.3</td>
</tr>
<tr>
<td>Water</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>17</td>
<td>2</td>
<td>0</td>
<td>20</td>
<td>0.9</td>
<td>0.15</td>
</tr>
<tr>
<td>Forest</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>15</td>
<td>2</td>
<td>20</td>
<td>0.8</td>
<td>0.25</td>
</tr>
<tr>
<td>None-Forest/None-Oil Palm</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>11</td>
<td>20</td>
<td>0.6</td>
<td>0.45</td>
</tr>
<tr>
<td>Total</td>
<td>25</td>
<td>10</td>
<td>13</td>
<td>18</td>
<td>20</td>
<td>29</td>
<td>25</td>
<td>140</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

User Accuracy 0.2 0.2 0.4 0.8 0.9 0.5 0.4 Overall Accuracy=50%
Commission 0.76 0.80 0.62 0.22 0.15 0.48 0.56 Kappa Co.=0.41
Table 3- Area covered by Oil palm plantation based on Landsat and PALSAR classification in Malaysia and Indonesia.

<table>
<thead>
<tr>
<th>Year</th>
<th>Area covered by Oil palm plantation (km²)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Indonesia</td>
<td></td>
</tr>
<tr>
<td>2017</td>
<td>297226</td>
</tr>
<tr>
<td>2015</td>
<td>348339</td>
</tr>
<tr>
<td>2010</td>
<td>334074</td>
</tr>
<tr>
<td>2008</td>
<td>326727</td>
</tr>
<tr>
<td>2005</td>
<td>164986</td>
</tr>
<tr>
<td>2000</td>
<td>161619</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Malaysia</td>
<td></td>
</tr>
<tr>
<td>2017</td>
<td>64743</td>
</tr>
<tr>
<td>2015</td>
<td>68632</td>
</tr>
<tr>
<td>2010</td>
<td>49343</td>
</tr>
<tr>
<td>2008</td>
<td>38573</td>
</tr>
<tr>
<td>2005</td>
<td>20265</td>
</tr>
<tr>
<td>2000</td>
<td>15265</td>
</tr>
</tbody>
</table>

Figure 5- Area covered by Oil palm plantation based on Landsat and PALSAR classification in Malaysia and Indonesia. a) Area of Oil palm plantation in Malaysia. b) Area of Oil palm plantation in Indonesia.
4. CONCLUSION

Our results show that from 2000 to 2017, the area covered by oil palm has drastically increased. However, it seems that the trend of expanding area is different in both countries (Figure 4). Malaysia has experienced a smoother trend of expanding the plantation while in Indonesia the plantation has been expanded differently. Nevertheless, some miss-classification may make the trend not continuously increasing (as it is expected), which is expected to overcome to this issue through increasing the sampling number from different year and region.

Our results show the potential use of GEE’s imagery classification system as a tool for oil palm land cover mapping but also reveal the limitations of this classification system especially in relation to the level of accuracy for detecting immature and mature oil palm plantations from other land cover types with similar spectral signatures. In most oil palm mapping studies, manual digitization of satellite imagery, accompanied by intensive field visits are commonly employed to detect oil palm from other land cover types (Carlson et al., 2013; Uryu et al., 2008). Such techniques ensure a higher level of accuracy and are able to differentiate immature, young plantations from other land cover types such as shrub or agricultural land. However, such high-level accuracy mapping techniques also require substantial expertise, resources and time, which is difficult to do on a frequent basis. Hence, there is a trade-off between time and resources, and the level of accuracy of oil palm mapping within GEE’s imagery classification system. The oil palm classification method demonstrated in GEE is useful to provide a quick understanding of oil palm plantations present in the landscape. This in itself is advantageous for independent monitoring bodies to conduct a survey of the landscape in question and conduct more detailed assessments if necessary.

In this study, we investigated the potential of GEE using Landsat and PALSAR to map oil palm plantations. Different from previous works, this study focused on country level classification with more complexity of landscapes. Our results showed that using the Landsat and PALSAR combination is better than using either Landsat or PALSAR alone. Many confusions between oil palm with other types (e.g. vegetation and forest) could be eliminated by using Landsat and PALSAR. It is a very promising approach to using PALSAR-2 (which is carried on board ALOS-2 launched in 2014) to provide better (higher resolution) observation on oil palm plantation.
References


VEGETATION CANOPY CHANGES AND USE OF SOILGRIDS DATA FOR ASSESSING THE EFFECT OF EXTREME RAINS ON ANNUAL SOIL LOSSES

Dhruba Pikha Shrestha
University of Twente
Faculty of Geo-information Science and Earth Observation (ITC)
P. O. Box 217, 7500 AE Enschede, The Netherlands
Email: d.b.p.shrestha@utwente.nl

KEY WORDS: Daily erosion model, vegetation canopy change modelling, rain interception, extreme rain, soil water balance

ABSTRACT

Extreme rains can trigger natural hazard processes such as soil erosion, land sliding and flash floods. Climate change studies show that the frequency of extreme rains is in an increasing trend, resulting in the amplification of hazard processes. For assessing the magnitude of soil losses various models are available. While annual empirical models (e.g. USLE, RUSLE, MMF) are easy to use, they do not take into account the effect of extreme rains. The event based models (e.g. LISEM, WEPP) can simulate erosion processes in detail, but rainfall event data is simply not available everywhere. To solve this problem, Shrestha and Jetten, (2018) have developed a daily erosion model and demonstrated that the effect of extreme rains can be incorporated easily in annual estimates. For running the model, daily rainfall, vegetation canopy changes, topography and soil data are required. Daily vegetation canopy changes mapping is a challenge and soil data may not be available easily everywhere due to higher cost involved in soil survey. Recently, time series NDVI and SoilGrids data are available freely, solving data scarcity problem. But, we do not know how good is the data for hazard assessment. The objectives of the study are in assessing the effect of daily canopy cover changes on rain interception, and in the use of SoilGrids data for erosion estimation. The study area is located in Sehoul, Morocco. Time series NDVI data at 1 Km resolution was obtained from Vito, Belgium (http://free.vgt.vito.be), and resampled to 15 m and at daily time step. Similarly, SoilGrids data at 250 m resolution was downloaded from ISRIC, The Netherlands (https://soilgrids.org). Pedotransfer functions were used to generate soil parameters and the daily erosion model was applied to assess soil losses. The results show that vegetation canopy cover plays an important role in the magnitude of soil losses. Canopy cover intercepts rain and protects the soil from raindrop impact. When canopy cover is lower, erosion rates are higher. During extreme rains, erosion can be very severe. The study shows that SoilGrids is a useful data source, and can be applied in daily erosion assessment in the semi-arid environment. The results also show that daily erosion modelling gives better picture of annual soil losses since the effects of extreme rains are also incorporated.

1. INTRODUCTION

Soil erosion is a wide spread problem occurring in all the climatic regions, which is responsible for decreased soil productivity endangering food security and causing offsite effects. Erosion rates can be alarming when excessive rain falls in a period when vegetation cover is lower (Vrieling et al., 2014). Extreme rains are often major causes of increased runoff, excessive soil losses, flash floods, and unwanted sedimentation problem in reservoirs, infrastructure and agriculture fields. Research on climate change shows the increase of the frequency of extreme rains (Board on Atmospheric Sciences and Climate, 2016), which demands our attention in understanding the associated problem so that preparations can be made in time in reducing the damage. For this, it becomes necessary to make first an assessment of the magnitude of erosion problem caused by extreme rains.

Various empirical models are available to assess soil losses e.g. RUSLE (Renard et al., 1991), MMF (Morgan, 2001), SLEMSA (Stocking, 1981) and SWAT (Gassman et al., 2007), but they are not capable of assessing the effects of extreme rains. Since the annual models use the average yearly rain and vegetation cover estimates, they ignore the effects of extreme rains. The event based models simulate soil detachment and transportation during a storm event e.g. LISEM (Jetten, 2002)(De Roo, 1996), WEPP (Laflen et al., 1991), EUROSEM (Morgan, et al., 1998), but detail rainfall data is almost impossible to get to run these models. Fortunately, assessing the effects of extreme rains is now possible using the recently available daily erosion model (Shrestha and Jetten, 2018).

The daily erosion model assesses rain interception loss to estimate effective rain for calculating soil losses. For this, it is necessary to map land cover and model vegetation canopy cover changes on a daily basis. The availability of time series NDVI data such as SPOT GVT and MODIS NDVI allow us to have temporal changes in vegetation cover.
Similarly, the daily erosion model estimates soil detachment by raindrop impact as well as by surface runoff. Surface runoff, estimated from soil water balance, is used for the transportation of detached soil particles. Soil parameters such as particle size distribution, water holding capacity of soil, hydraulic conductivity and porosity are required in modelling water fluxes in the soil. But, detail soil data is often not available easily because of high costs involved making surveys. Recently, 250m SoilGrids data, based on global interpolation and machine learning techniques on 150,000 soil profiles data collected worldwide, is available freely (Hengl et al., 2017). The SoilGrids data solves the soil data scarcity problem but its usefulness and application on erosion modelling has not been yet done. The objective of the research is on using time series NDVI data for assessing daily rain interception and the usefulness of SoilGrids data for assessing the effects of extreme rain. It is applied in a case study in Sehoul commune in Morocco.

2. STUDY AREA

The study area is located in between 6° 34’ and 6° 46’ W longitudes, 33° 52’ and 33° 59’ N latitudes, covering approx. 96 sq.km and about 20 km east of Rabat in Morocco (Figure 1). Elevation varies from 45 to 233 m asl. The area receives in average 540 mm rain (Rabat/Sale meteorological station) calculated over a 59 year period (1951-2010). There is inter-annual variation of precipitation with some years receiving below average rains causing extended drought while some years receive above average rains. The river Bou Regreg drains the area. The main land cover/land use types are cork oak forest, plantation forest, rainfed agriculture, and grazing land. Grazing takes place in the cork forest since it is not protected. Insufficient precipitation has been the common problem for cultivation. Low vegetation cover and intense rain results in excessive soil losses. Extensive gully formation occurs not only in grazing land but also in the forest (DESIRE, 2013). When gully formation is severe, cultivation is not economical and the land is often abandoned making the situation very severe.

Figure 1. Study area

3. MATERIALS AND METHODS

3.1 Generation of land cover map

Aster multi-spectral data (ASTL1 B data), acquired on 14 July 2003 at spatial resolution of 15 m, was obtained. The data was pre-processed for atmospheric correction before running maximum likelihood classifier. Sufficient points were collected to be used as training samples for classification as well as for accuracy assessment.

3.2 Generation of daily canopy cover maps

For erosion studies it is essential to estimate rain interception by vegetation canopy. The effective rain for splash detachment and for runoff estimation is total rain minus rain interception. Interception storage is estimated by computing leaf area index (LAI) which depends on vegetation type and is generally derived from canopy cover. SPOT VGT images at 1 km spatial and 10 days resolutions, consisting of total 38 images were obtained from Flemish Institute for Technological Research, Belgium (http://free.vgt.vito.be). The images were resampled to spatial resolution of 15m. The obtained time series SPOT VGT data were transferred to NDVI values ranging from -1 to 1 to make it easier to interpret as follows:

\[ NDVI = DN \times 0.004 - 0.1 \] (1)
From the obtained time series data, linear interpolation was carried out to obtain daily NDVI images as follows:

\[ y = y_1 + \frac{x-x_1}{x_2-x_1} (y_2 - y_1) \]  

(2)

Where, \( y \) is pixel value at location \( x \), between two values \( y_1 \) and \( y_2 \) at positions \( x_1 \) and \( x_2 \). The \( x_1 \) and \( x_2 \) are the NDVI images on 2 dates (day number in the year) and \( y_1 \) and \( y_2 \) are the corresponding NDVI values. From the NDVI the vegetation cover factor is generated using the exponential function (Van der Knijff et al., 1999) as follows:

\[ \text{Cover} = 1 - e^{-a \frac{(NDVI)}{(\beta - NDVI)}} \]  

(3)

Suggested values for the constants \( a \), \( \beta \) are 2 and 1.

3.3 SoilGrids data and pedotransfer functions

SoilGrids data layers of particle size distribution (sand, silt and clay percentages), organic matter content, coarse fragments and bulk density of topsoil (15 cm) were downloaded from ISRIC, The Netherlands (https://soilgrids.org). The obtained data in 250 m resolution was resampled to 15 m. Pedotransfer functions were used to generate soil parameters required to run the daily erosion model. The parameters are saturated hydraulic conductivity (mm/hr), field capacity and wilting point according to (Saxton and Rawls, 2006), soil porosity using bulk density and particle density and soil erodibility factor based on (Wischmeier and Smith, 1960). Soil cohesion was estimated based on clay content, which is adapted from (Morgan, 2001). The codes for pedotransfer functions, written in PCRaster, were then applied to SoilGrids data.

3.4 Topographic data

For topographic data analysis, SRTM DEM at 30 m resolution obtained from US Geological Survey’s EROS data center (https://earthexplorer.usgs.gov) was resampled to 15 m resolution. From this, slope gradient map was generated which was used in the estimation of transport capacity of overland flow, and flow network (Figure. 2) used in routing surface runoff. The flow network was generated using the direction of maximum slope gradient in a grid cell.

Figure 2. Generation of flow network for routing surface runoff based on maximum slope direction in a grid cell.

3.5 Daily erosion modelling

The daily erosion model estimates rain interception as well as soil water balance on a daily basis. Soil detachment and runoff estimation is based on effective rain (daily rain minus canopy interception) multiplied by a runoff fraction, which is based on soil water balance. Detail description of the model can be found at (Shrestha and Jetten, 2018). It is outside the scope of this paper to describe in detail how the daily erosion model works. The model codes were written in PCRaster, an open source GIS software for environmental modelling (http://pcraster.geo.uu.nl). The parameters required to run the model are: daily rainfall (mm), daily vegetation canopy cover (in percentage), soil parameters (soil porosity in percentage, saturated hydraulic conductivity in mm/hr, soil erodibility factor in 0 to 1, cohesion in kPa) and topography (slope gradient in percentage, flow network). Detail model parameters are given in Table 1. The model gives results on daily soil detachment, erosion and deposition (t/ha). The daily erosion estimates are added together to get annual soil loss. The results can be assessed per land cover/land use type. The daily erosion model was run twice: using SoilGrids data and using soil data collected from the field. In addition, the model was also run using 2 rainfall data: one from rainfall (2003) which represents long term average rain in Morocco, and one from the year with extreme rainy days (1975).
Table 1. Input parameters for the Daily erosion model

<table>
<thead>
<tr>
<th>Factor</th>
<th>Parameter</th>
<th>Definition and remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Climate</td>
<td>rain</td>
<td>Daily rainfall (mm)</td>
</tr>
<tr>
<td></td>
<td>ETo</td>
<td>Potential evapotranspiration on a daily basis (mm)</td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>Rainfall intensity (mm/hr: 10 for temperate climate, 25 for tropical and 30 for season climates (e.g. Monsoon or Mediterranean)</td>
</tr>
<tr>
<td>Soil</td>
<td>theta_s</td>
<td>Porosity (volumetric percentage)</td>
</tr>
<tr>
<td></td>
<td>theta_fc</td>
<td>Field capacity (volumetric %)</td>
</tr>
<tr>
<td></td>
<td>Theta_wp</td>
<td>Wilting point (volumetric %)</td>
</tr>
<tr>
<td></td>
<td>thetainit</td>
<td>Initial soil moisture (volumetric percentage, e.g. 0.5 of wilting point</td>
</tr>
<tr>
<td></td>
<td>coh</td>
<td>Cohesion (kPa)</td>
</tr>
<tr>
<td></td>
<td>K</td>
<td>Soil erodibility factor (0-1)</td>
</tr>
<tr>
<td></td>
<td>ksat</td>
<td>Saturated hydraulic conductivity of soil (mm/hr)</td>
</tr>
<tr>
<td>Topography</td>
<td>DEM</td>
<td>Digital elevation model</td>
</tr>
<tr>
<td></td>
<td>grad</td>
<td>Slope gradient (%)</td>
</tr>
<tr>
<td></td>
<td>ldd</td>
<td>Local drainage direction map</td>
</tr>
<tr>
<td>Land cover</td>
<td>landuse</td>
<td>Land cover/land use map</td>
</tr>
<tr>
<td></td>
<td>NDVI</td>
<td>Time series NDVI maps (e.g. VGT SPOT)</td>
</tr>
<tr>
<td></td>
<td>Kc</td>
<td>FAO crop factor</td>
</tr>
<tr>
<td></td>
<td>Cover</td>
<td>Canopy cover (% expressed in 0-1)</td>
</tr>
<tr>
<td></td>
<td>GC</td>
<td>Ground cover (% expressed in 0-1)</td>
</tr>
<tr>
<td></td>
<td>PH</td>
<td>Plant height (m)</td>
</tr>
</tbody>
</table>

4 RESULTS AND DISCUSSIONS

4.1 Land cover and rain interception

Land cover classification result is shown in Figure 3. Overall classification accuracy obtained is 89%. The area has dominantly cropland (3671 ha), which is followed by grazing land (2826 ha), cork forest (1818 ha), bare soil (1103 ha) and plantation forest (200 ha). In terms of obtaining time series NDVI images, it was not a problem to get cloud free images for Morocco since it is located in drier climatic region. Linear interpolation of 10 day NDVI images into daily estimates shows good response of daily rainfall to vegetation cover. Vegetation cover decreases as from March to the minimum NDVI values (less than 0.4) during July/August and it increases as from October when it starts to rain until December when it reaches maximum (NDVI value of about 0.7).

4.2. Soil loss estimates

Soil parameters for running daily erosion model were derived by applying pedotransfer functions on the SoilGrids data. The result is shown in Table 2 and in Figure 4. Highest soil losses (>50 t/ha) were recorded in the cork forest and in the grazing land, which are located in the steep slopes (>30% slopes). Because of over grazing no undergrowth of vegetation is possible in the forest areas. The reason for high soil losses in the plantation forest could be due to the young age of the trees, with low canopy cover and lower soil protection. Lower soil losses (4-6 t/ha) were recorded in the gently sloping areas for all the land cover types. Erosion assessment was also carried out using soil parameters collected in the field and from laboratory analysis of soil samples. The results show that there is not considerable variation in soil loss assessments whether field based soil data or SoilGrids data is used (Table 2). Some differences were found in plantation forest and in grazing land, especially in the steeper slopes (> 30% slopes). In general, soil losses were found to be relatively higher if SoilGrids data is used but the differences are not so big. It seems that soil loss estimation is limited by the transport capacity of the runoff water. Soil detachment can be higher but it seems that everything cannot be transported down the hill and is limited by the capacity of the runoff water. Soil erodibility factor in field based data varies from 0.1 to 0.9 whereas it varies from 0.2 to 0.4 in SoilGrids data, which means higher soil detachment rates if field data is used instead of SoilGrids data. Similarly, saturated hydraulic conductivity varies from 1.3 to 20 mm/hr for field data whereas it varies from 0.9 to 154 mm/hr in case of SoilGrids data. This shows that rain infiltration will be relatively higher in case of SoilGrids data. In both the cases runoff generation and the transportation capacity seems to be somewhat similar. Although soil detachment rate is higher transport capacity seems to be the one which determined the soil losses. This could be the main reason why soil loss estimates are similar in both cases: using fieldwork based soil data or using SoilGrids data.
Figure 3. Land cover classification

Table 2. Soil loss assessment using field data and SoilGrids data

<table>
<thead>
<tr>
<th>Land cover</th>
<th>Slope classes</th>
<th>Surface area</th>
<th>Soil loss assessment with field data</th>
<th>Soil loss assessment with SoilGrids data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Hectare</td>
<td>Percent</td>
<td>t.ha⁻¹</td>
</tr>
<tr>
<td>Agriculture</td>
<td>Level to gently sloping (0-8%)</td>
<td>1990</td>
<td>54</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Rolling to hilly (8-30%)</td>
<td>1574</td>
<td>43</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>Steeply dissected to mountainous (&gt;30%)</td>
<td>107</td>
<td>3</td>
<td>31</td>
</tr>
<tr>
<td>Bare soil</td>
<td>Level to gently sloping (0-8%)</td>
<td>663</td>
<td>60</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Rolling to hilly (8-30%)</td>
<td>426</td>
<td>39</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>Steeply dissected to mountainous (&gt;30%)</td>
<td>14</td>
<td>1</td>
<td>27</td>
</tr>
<tr>
<td>Cork forest</td>
<td>Level to gently sloping (0-8%)</td>
<td>1257</td>
<td>69</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Rolling to hilly (8-30%)</td>
<td>483</td>
<td>27</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>Steeply dissected to mountainous (&gt;30%)</td>
<td>78</td>
<td>4</td>
<td>52</td>
</tr>
<tr>
<td>Plantation forest</td>
<td>Level to gently sloping (0-8%)</td>
<td>143</td>
<td>72</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Rolling to hilly (8-30%)</td>
<td>30</td>
<td>15</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>Steeply dissected to mountainous (&gt;30%)</td>
<td>27</td>
<td>13</td>
<td>34</td>
</tr>
<tr>
<td>Grazing land</td>
<td>Level to gently sloping (0-8%)</td>
<td>808</td>
<td>29</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Rolling to hilly (8-30%)</td>
<td>1512</td>
<td>53</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>Steeply dissected to mountainous (&gt;30%)</td>
<td>506</td>
<td>18</td>
<td>47</td>
</tr>
</tbody>
</table>
The result also shows that cork forest intercepts the highest rainfall (140 mm rain which is ¼ of the annual rain) and lowest interception is by plantation forest (57 mm) (Table 3). The reason for having lower interception in plantation forest could be the young age of the plantation trees. Higher rain interception in grazing land is probably due to the presence of shrubs.

### Table 3. Total interception in different land cover types in 2003

<table>
<thead>
<tr>
<th>Land cover</th>
<th>Surface area (ha)</th>
<th>Total interception mm/(percent of annual rain)</th>
<th>Annual rain mm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agriculture</td>
<td>3671</td>
<td>78 (0.14)</td>
<td>568</td>
</tr>
<tr>
<td>Bare soil</td>
<td>1103</td>
<td>71 (0.12)</td>
<td>568</td>
</tr>
<tr>
<td>Cork forest</td>
<td>1818</td>
<td>140 (0.25)</td>
<td>568</td>
</tr>
<tr>
<td>Plantation forest</td>
<td>200</td>
<td>57 (0.10)</td>
<td>568</td>
</tr>
<tr>
<td>Grazing land</td>
<td>2826</td>
<td>117 (0.20)</td>
<td>568</td>
</tr>
</tbody>
</table>

The model was run again with rainfall data from 1975 to see the effect of extreme rainy days. Total rain in 2003 was 658 mm and in 1975 annual rain was 669 mm, both year receiving similar amount of rain. In 1975 there were few days with extreme rain (40 mm) with the maximum amount received in 24 hour being 149 mm on 17 Dec, which accounts for about one fifth of the annual rain (Table 4). Soil loss results also show the effect of the presence of extreme rainy days in 1975. Although annual rain in both the years can be considered similar (658 mm rain in 2003 and 669 mm rain in 1975) soil losses in 1975 are almost double in all the land cover types, which is mainly due to the presence of extreme rains (Figure 5). In 2003 only 2 days received rain above 40 mm rain.

### Table 4. Normal rain in 2003 and the year with extreme rainy days (1975)

<table>
<thead>
<tr>
<th>Land cover</th>
<th>2003</th>
<th>1975</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Annual rain 658 mm (max 24 hr rain 72 mm)</td>
<td>Annual rain 669 mm (max 24 hr rain 149 mm)</td>
</tr>
<tr>
<td></td>
<td>Soil loss annual (t/ha)</td>
<td>Soil loss 9 Dec 2003</td>
</tr>
<tr>
<td>Agriculture</td>
<td>12</td>
<td>5</td>
</tr>
<tr>
<td>Bare soil</td>
<td>10</td>
<td>4</td>
</tr>
<tr>
<td>Cork forest</td>
<td>14</td>
<td>7</td>
</tr>
<tr>
<td>Plantation forest</td>
<td>18</td>
<td>8</td>
</tr>
<tr>
<td>Grazing land</td>
<td>20</td>
<td>10</td>
</tr>
</tbody>
</table>
4. CONCLUSION

The result shows that rain interception can be higher in drier regions (upto 25% of the annual rain) depending on vegetation types. The maximum annual interception of 140 mm was estimated for the cork forest. This is due to the dryness of the area. The result also shows that SoilGrids data can be applied easily in erosion studies in combination with modelling vegetation canopy cover changes. It seems that soil loss is limited by the transport capacity of the runoff water in the semi-arid environment. The results also shows that daily erosion modelling gives better picture of soil losses since it also incorporates the effects of extreme rains, which is not possible by using annual erosion models. The daily erosion model can be applied using SoilGrids data in areas with similar landscape and environmental conditions like in Morocco. For application of the method in other areas, especially with different climatic conditions and in varying geomorphic setting, the method will have to be tested first.

ACKNOWLEDMENTS

ASTER and Shuttle Radar Topographic Mission (SRTM) data were downloaded from the US Geological Survey’s EROS data center. Time series NDVI images were obtained from the Flemish Institute for Technological Research, Belgium. SOILGRIDS data was obtained from ISRIC, The Netherlands. Fieldwork was financed by the EU DESIRE project (http://www.desire-project.eu).

REFERENCES