Spectral discrimination of healthy and Ganoderma-infected oil palms from hyperspectral data

Helmi Z. M. Shafri, Mohd I. Anuar, Idris A. Seman & Nisfariza M. Noor

Geomatics Engineering Unit, Department of Civil Engineering, Universiti Putra Malaysia (UPM), 43400, Serdang, Selangor, Malaysia

Malaysian Palm Oil Board (MPOB), No. 6, Persiaran Institusi, Bandar Baru Bangi, 43000, Kajang, Selangor, Malaysia

Department of Geography, Faculty of Arts and Social Sciences, University of Malaya (UM), 50603, Kuala Lumpur, Malaysia

Available online: 02 Aug 2011


To link to this article: http://dx.doi.org/10.1080/01431161.2010.519003

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.tandfonline.com/page/terms-and-conditions

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings,
demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.
Spectral discrimination of healthy and *Ganoderma*-infected oil palms from hyperspectral data

HELMI Z. M. SHAFRI*,†, MOHD I. ANUAR‡, IDRIS A. SEMAN‡ and NISFARIZA M. NOOR§

†Geomatics Engineering Unit, Department of Civil Engineering, Universiti Putra Malaysia (UPM), 43400 Serdang, Selangor, Malaysia
‡Malaysian Palm Oil Board (MPOB), No. 6, Persiaran Institusi, Bandar Baru Bangi, 43000 Kajang, Selangor, Malaysia
§Department of Geography, Faculty of Arts and Social Sciences, University of Malaya (UM), 50603 Kuala Lumpur, Malaysia

*(Received 26 September 2008; in final form 12 July 2010)*

Although hyperspectral remote sensing has been used to study many agricultural phenomena such as crop stress and diseases, the potential use of this technique for detecting *Ganoderma* disease infestations and damage to oil palms under field conditions has not been explored to date. This research was conducted to investigate the feasibility of using a portable hyperspectral remote-sensing instrument to identify spectral differences between oil-palm leaves with and without *Ganoderma* infections. Reflectance spectra of samples representative of three classes of disease severity were collected. The most significant bands for spectral discrimination were selected from reflectance spectra and first derivatives of reflectance spectra. The significant wavelengths were identified using one-way analysis of variance. Then, a Jeffries–Matusita (JM) distance measurement was used to determine spectral separability between the classes. A maximum likelihood classifier method was used to classify the three classes based on the most significant wavelength spectral responses, and an error matrix was finally used to assess the accuracy of the classification.

1. Introduction

Oil palm is a commercial crop that is important in food product manufacturing and is a source of biomass fuels. Oil-palm trees are grown in fertile, tropical environments. They produce fruit that can be processed into low-cholesterol cooking oils and can be used as one of the ingredients in environmentally friendly biofuels that reduce engine emissions (Corley and Tinker 2003). However, oil-palm plantations in Malaysia are facing problems associated with *Ganoderma* disease infections that reduce the quantity and quality of the oil-palm products and that have resulted in economic losses (Miller *et al.* 1999, Flood *et al.* 2000).

*Ganoderma*, also known as white-rot fungus, is an organism that causes damage to the oil palm and can lead to economic hardship. The term ‘white rot’ is derived from the fungus that affects the lignin within the wood and exposes white cellulose
The Ganoderma fungus commonly spreads via spores and grows in the non-living features of oil palms (Flood et al. 2000, Paterson et al. 2000). In this article, we focus on studying Ganoderma infections in the seedlings of oil-palm trees. For oil-palm seedlings, the symptoms of Ganoderma are yellowing and desiccation (browning) of leaves, from the oldest to the youngest fronds. Fruiting bodies may or may not have developed on the infected seedlings before or after the appearance of the foliar symptoms. Internal disease symptoms were obvious on seedlings inoculated with wood blocks treated with Ganoderma pathogens (Idris et al. 2006). Hyperspectral remote sensing, with its abundance of spectral data, seems more useful than a multispectral sensor for the detailed mapping or identification of vegetation properties (Mutanga et al. 2003, Zarco-Tejada et al. 2004, Jensen 2005). The motivation to use higher dimensionality data is that doing so helps with the discrimination of much larger numbers of features and classes that are more detailed in nature. However, the wealth of data makes analysis more complex, and there is a need for more suitable signal models to process the data (Landgrebe 2003). The applications of hyperspectral remote-sensing data are broad, including species discrimination, nutrient content assessment, stress detection and disease detection (Schmidt and Skidmore 2002, Lelong et al. 2006, Mirik et al. 2006, Delalieux et al. 2007).

The narrow bandwidth of hyperspectral data can be used to measure more specific spectral vegetation properties (Ismail et al. 2008). In hyperspectral remote-sensing studies, methods such as Red Edge Position (REP) are widely used to assess stress, water content, nitrogen content and chlorophyll content of vegetation (Dawson and Curran 1998, Liu et al. 2004, Baranoski and Rokne 2005, Cho et al. 2005). However, this method only focuses on the red-edge region (680–780 nm). The red-edge region is known as the region of abrupt change in leaf reflectance between 680 and 780 nm caused by combined effects of strong chlorophyll absorption in the red wavelengths and high reflectance in the near-infrared wavelengths due to leaf internal scattering (Cho and Skidmore 2006).

There are other non-REP methods that use hyperspectral remote-sensing data to differentiate between vegetation types. These methods make use of band selection to identify wavelengths that are significant for disease detection and vegetation species discrimination. This research uses a partial least square discriminant analysis (PLS-DA), parametric or non-parametric discriminative statistical analysis, standard deviations, stepwise discriminant analysis, principal component analysis (PCA) and genetic algorithms (GAs) to identify wavebands that reveal significant differences between different vegetation types or within vegetation types (Osborne et al. 1997, Schmidt and Skidmore 2002, Fung et al. 2003, Zhang et al. 2003, Li et al. 2007, Vaiphasa et al. 2007).

Several reports to date have focused on using the raw reflectance spectra and also on using the derivatives of reflectance spectra to analyse the responses of these spectra to vegetation or other features that are of relevance for vegetation studies (Malthus and Madeira 1993, Post et al. 2007, Zhan-Yu et al. 2008). The purpose of using derivatives in this context is to eliminate background signals and resolve overlapping spectral features. Derivative techniques can also better focus on the most important information and eliminate any unnecessary information, and they can highlight spectral features that exhibit sharp structures (Demetriades-Shah et al. 1990, Smith et al. 2004). In this study, the first derivative of the reflectance spectra was analysed to determine whether the derivative technique can improve the ability to detect the health of oil-palm trees.
The high number of significant wavelengths in vegetation discrimination led to the use of a separability index such as the Jeffries–Matusita (JM) distance index (Schmidt and Skidmore 2002, Vaiphasa et al. 2005, Ismail et al. 2008, Jollineau and Howarth 2008). This separability index is useful for filtering out highly correlated wavelengths that will not provide a reliable classification solution (Wang and Sousa 2009). The JM distance was used to predict the potential for the correct classification of salt marsh vegetation types as studied in Schmidt and Skidmore (2002). Jollineau and Howarth (2008) used the JM distance to quantify the separability distance of the most significant wavelength in the maximum likelihood classification of inland wetland complexes.

Vaiphasa et al. (2005) also studied the use of four preferred wavelengths as inputs in calculating the JM spectral distance between 16 tropical mangrove species. The results indicate that the best wavelength combinations gave an acceptable JM distance for almost all the mangrove species except for the Rhizophoraceae family of mangroves. The JM distance was also used by Ismail et al. (2008) to quantify the distance given by the best wavelength combination for identifying Sirex noctilio in pine forest plantations in South Africa. The results show that the combination of more than seven preferred wavelengths yielded an acceptable JM distance to discriminate between healthy and diseased pine trees.

Despite the success of hyperspectral remote-sensing techniques in detecting other types of agricultural crop diseases, the feasibility of detecting Ganoderma in oil palms has not been assessed to date. Research on oil palms has only been conducted to study nutrition deficiencies using hyperspectral data (Lelong et al. 2006); the conclusion was that hyperspectral remote-sensing data can offer significant potential for these types of measurements.

This study was conducted to identify wavelengths from two types of data (reflectance spectra and the first derivatives of the reflectance spectra) to assist in discriminating between healthy and Ganoderma-infected oil palms. The results obtained from this study may be used to form the basis of future algorithms or spectral indices for early disease detection using airborne or space-based hyperspectral remote-sensing platforms. Specifically, the objectives of this study are (1) to determine whether significant differences are exhibited by the reflectance spectra and by the first derivatives of the reflectance spectra between the healthy and infected classes; and (2) to identify the most significant wavelengths from reflectance spectra together with the first derivative of the reflectance spectra in order to discriminate between the two classes.

2. Material and methods

The general flow chart of our methodology is shown in figure 1.

2.1 Study area and data collection

The data used in this study were acquired in December 2007 from an oil-palm nursery managed by the Malaysian Palm Oil Board (MPOB) located in Bangi, Selangor, Malaysia (2°54′24.30″ N, 101°47′18.36″ E). The oil-palm tree samples consisted of 102 germinated seeds. Of those, 68 samples were planted in polypropylene bags containing rubberwood blocks that had been inoculated with Ganoderma pathogens.

The samples were 4 months old during the inoculation, and they were arranged in two plots. The first plot contained 34 control (healthy) samples, while the second
contained 68 samples that had been inoculated with *Ganoderma* pathogens. The trees were maintained with regular watering, manure and pesticide applications. The infection with *Ganoderma* and associated disease developments were assessed monthly during a period of 6 months to observe foliar symptoms visually including white mycelium, progressive yellowing, desiccation (browning) of the oldest to the youngest fronds and death of the seedlings with or without *Ganoderma* fructifications.

After 6 months, the samples from the second plot were separated into mild and severe symptoms. The differences between the disease levels in the second plot were due to the resistance of the samples to the *Ganoderma* pathogen. The mild and severe disease samples were categorized by oil-palm disease specialists from MPOB. The samples were segregated into three classes designated T1, T2 and T3. Each class included 34 samples. Class T1 represented healthy samples, T2 represented the samples with mild symptoms of *Ganoderma* infection and T3 represented samples with severe *Ganoderma* infections. Mildly diseased plants did not show any foliar symptoms, white mycelium or fruiting bodies, but there was root damage. In the case of severe infections, there were fruiting bodies and white mycelia at the leaves and trunks of the samples, together with evident yellowing.

During our fieldwork efforts, the reflectance spectra of the oil-palm leaves were collected using an APOGEE spectroradiometer instrument with a spectral range of 350–1000 nm, a 30° field of view and a spectral resolution of 0.5 nm. The field samples were taken on a sunny clear day from 10:00 am to 2:00 pm local time. Calibration was performed using a white reference (barium sulphate, BaSO₄) before taking each reading. Six readings of reflectance spectra were taken from two leaflets of each sample. The readings were taken from the top first and second leaflets, which were young fronds, as shown in figure 2. The reflectance spectra readings were averaged to yield mean reflectance spectra for individual samples.
Figure 2. Locations of first and second fronds for reflectance spectra acquisition.

Figure 3 shows the visual interpretation of T1, T2 and T3. There are no obvious differences between T1 and T2 upon visual inspection. For T3, the yellowing effect is very apparent, which indicates a severe *Ganoderma* infection. *Ganoderma*'s impacts on sample roots can be seen in figure 4(a) and (b).

2.2 Preprocessing

The data indicate the presence of random noise, which is made apparent by taking the first derivative of the reflectance spectra, before any denoizing techniques are applied to the raw data set. Each of the 34 reflectance spectra from each class was first denoized using wavelet analysis, according to the guidelines suggested by Schmidt and Skidmore (2004). The mother wavelet used for denoizing the raw reflectance spectra was Symlet 8 with two levels of decomposition and zeroed detail coefficients.

The denoized spectral reflectance data were then transformed to obtain the first derivative of the reflectance spectra. Only the original reflectance spectra at 460–959 nm were used in this study because all the other ranges contained very pronounced noise that could not be cleaned or smoothed. The denoized reflectance spectra were then transformed into first derivatives of the reflectance spectra using

$$\text{FD} = \frac{y_2 - y_1}{x_2 - x_1}$$

(1)

where FD is the first derivative that was computed with inputs $y_2$ and $y_1$ – namely, the reflectance spectra at the second and first wavelengths. The values $x_2$ and $x_1$ represent the second and first wavelengths.

2.3 Statistical test

A statistical test was used to compare the responses associated with the reflectance spectra and to compare the first derivatives of the reflectance spectra of three different health levels of oil-palm trees that were infected by *Ganoderma*. This allowed us to examine whether at least one pair of categories was statistically different at every spectral band. For the purposes of this test, the null hypothesis was $H_0: \mu_1 = \mu_2, \mu_3$ versus the alternative hypothesis where $H_a: \mu_1 \neq \mu_2, \mu_3$. In this case, $\mu_i$ were the mean
spectra values of the $i$th class ($i = 1, 2, 3$). The distribution of spectral responses at every spectral wavelength was assumed to be normal, consistent with the central limit theorem (number of spectra, $N \geq 30$) (Vaiphasa et al. 2005, Wang and Sousa 2009). We tested our hypothesis using a one-way analysis of variance (ANOVA) at every wavelength between 460 and 959 nm (a total of 999 wavelengths) with 95% confidence limits ($\alpha = 0.05$).
Figure 4. (a) Symptoms of *Ganoderma boninense*. Uninoculated samples without disease symptoms (1A) and inoculated samples with foliar symptoms and *Ganoderma* fruiting bodies at 6 months after inoculation (1B). (b) Inoculated with internal symptoms (2A) and uninoculated without internal symptoms (2B). Oil-palm samples at 6 months after inoculations (Idris *et al.* 2006).

A one-way ANOVA test was chosen to replace the direct graphical presentations of the oil-palm spectral responses because direct visualization was not an effective visualization tool for comparing the three health categories with the 34 spectral responses from reflectance spectra and the first derivatives of these reflectance spectra. The spectral variations within individual health categories led to spectral overlaps that made it very difficult to spot the differences between the three classes with the naked eye (Landgrebe 2003).

Other researchers have used the U test and the H test in order to identify significant wavelengths for vegetation discrimination applications (Schmidt and Skidmore 2002, Liu *et al.* 2005). However, both tests are valid only for non-parametric data where there are fewer than 30 reflectance spectra of each wavelength from each different class and where this information is not normally distributed.

The ANOVA test is recommended as an alternative to visual difference interpretation to identify those wavelengths that exhibit insignificant and significant spectral separability for normally distributed data sets. Poor wavelength locations are those for which $p$-values are $\geq 0.05$, and significant wavelengths will result in $p$-values that are $<0.05$. For wavelengths with $p$-values greater than the $\alpha$ threshold, we conclude that the spectra of each class are very similar and that there is no meaningful separability between data from different classes.

However, if the $\alpha$ value is $<0.05$, this indicates that the spectra are statistically separable in at least one pair of classes. Although this test can assist in rapidly identifying significant wavelengths, the test results cannot be independently interpreted without additional treatment because a type 1 error usually occurred when testing multiple hypotheses (Vaiphasa *et al.* 2005). Thus, the spectral separability index for every class pair must be calculated to guarantee the existence of a real difference between the spectra of the various health classes.
2.4 Spectral separability

The separability index used in this study relied on a JM distance analysis. The JM distance calculation yields a value between 0 and $\sqrt{2}$ ($\approx 1.4142$), with higher values representing better separability between class pairs (Richards and Jia 1999). Therefore, the band or band combination that produces the highest JM distance averaged over each class pair can be considered most suitable to discriminate between Ganoderma disease classifications. The usual threshold to indicate separability using JM is $\geq 95\%$ or $\geq \sqrt{1.9}$ (Vaiphasa et al. 2005). The JM distance can be formally expressed as

$$JM_{ij} = \sqrt{2 \left(1 - e^{-\alpha}\right)}$$

where $JM_{ij}$ represents the JM distance measure between class $i$ and $j$ and $e$ is the exponential for $\alpha$, which represents the Bhattacharyya distance, calculated using

$$\alpha = \frac{1}{8} (\mu_i - \mu_j)^T \left(\frac{C_i + C_j}{2}\right)^{-1} (\mu_i - \mu_j) + \frac{1}{2} \ln \left(\frac{|C_i + C_j|}{|C_i| |C_j|}\right)$$

where $i$ and $j$ are the two classes being compared; $C_i$ is the covariance matrix of data in class $i$; $\mu_i$ is the mean vector of data in class $i$; $\ln$ is the natural logarithm; and $|C_i|$ is the determinant of $C_i$ (matrix algebra).

The separability between class pairs was measured to quantify the separability of the most significant wavelengths that should be used as a classification input. Our goal was to select the most significant wavelengths with good separability such that spectral discrimination between classes can be maximized while omission and commission errors will be minimized (Bruzzone and Serpico 2000, Jensen 2005, Jollineau and Howarth 2008). This procedure can increase the likelihood of an accurate classification result.

2.5 Supervised classification

A simple supervised classification method was used to examine the accuracy of classifying the three classes. The maximum likelihood classifier method was selected as a means of examining the spectral classification accuracy of the most significant band combinations. This classifier assesses both variance and covariance when classifying unknown spectral responses. The maximum likelihood classifier is a parametric classifier that assumes the spectral responses of each class are normally distributed (Jollineau and Howarth 2008).

The maximum likelihood approach was chosen because it has been widely used in many studies involving hyperspectral data (Held et al. 2003, Clark et al. 2005, Salem et al. 2005, Chen et al. 2007, Gagnon et al. 2007, Lu et al. 2007, Jollineau and Howarth 2008, Yang et al. 2009). We tested our results for accuracy by using an error matrix (Foody 2002, Jensen 2005). In the error matrix, the overall accuracy and Kappa statistics can be calculated, and classifications are considered robust if the Kappa coefficient value $>0.8$ (80%). Kappa values of 0.4 to 0.8 (40% to 80%) represent moderate classification accuracy, and Kappa coefficient values $<0.4$ (40%) indicate poor classification accuracy (Jensen 2005).
3. Results

3.1 Preprocessing

During the preprocessing stage, we used wavelet denoising to reduce noise effects and to obtain smoother data without losing important information and features of the original reflectance spectra. Symlet 8 was used for data denoizing, as suggested by Schmidt and Skidmore (2004).

Figure 5(a) and (b) shows the effect of denoizing on the first derivative of reflectance spectra. The first derivative of the reflectance spectra enhanced the presence of noise in the data, as apparent from the pronounced peaks and valleys. The denoising procedure reduced the noise effects, as seen in figure 5(b).

3.2 Statistical testing

We wished to determine the most suitable wavelengths for spectral discrimination. The results of one-way ANOVA testing for individual class combinations (T1 vs. T2, T2 vs. T3 and T1 vs. T3) for reflectance spectra consistent with the first derivative of the reflectance spectra are shown in figures 6(a)–(c) and 7(a)–(c). The shaded areas indicate the spectral wavelength locations, while the class pairs exhibit statistically significant differences in spectral response ($p < 0.05$).

![Figure 5](image_url)  
Figure 5. Results of denoizing: (a) original first derivative of the reflectance spectra, (b) denoized first derivative of the reflectance spectra.
Figure 6(a)–(c) suggests that none of the reflectance spectra at any wavelength indicate significant differences between T1 and T2. The reflectance spectra of T1 and T2 are very similar for each wavelength, and visual inspection confirms that the data sets should not exhibit any obvious differences between classes. However, the reflectance spectra data sets show significant differences in several spectral regions.

Figure 6. ANOVA results for individual class pairs consistent with the reflectance spectra data set: (a) healthy (T1) versus mild infection (T2); (b) mild (T2) versus severe infection (T3); and (c) healthy (T1) versus severe infection (T3). The darker shading indicates regions of the electromagnetic spectrum that indicated significant differences.
Spectral discrimination of oil palm using hyperspectral data

Figure 7. ANOVA results for individual class pairs in the context of the first derivative of the reflectance spectra: (a) healthy versus mild infection; (b) mild infection versus severe infection; and (c) healthy versus severe infection. Dark shading indicates regions of the electromagnetic spectrum where there were significant differences.

for the T2 versus T3 and for the T1 versus T3 class pairs. The reflectance spectra exhibited significant differences in the blue (71 wavelengths), green (103 wavelengths), red (92 wavelengths) and infrared (276 wavelengths) regions for the T2 versus T3 class pair. T1 versus T3 differences are apparent in the blue (71 wavelengths), green
(150 wavelengths), red (104 wavelengths) and infrared (418 wavelengths) regions. Severely infected plants can thus be discriminated from others by using the visible and infrared bands.

Our ANOVA test results also indicate that there are a greater number of significant wavelengths that can be used to discriminate between T1 and T3 within each significant spectral region. However, the ability to distinguish between healthy (T1) and mild (T2) infections cannot be articulated using reflectance spectra data sets alone. Instead, we used the first derivative of the reflectance spectra to examine whether our approach could be improved.

ANOVA test results with the first derivative data sets (as shown in figure 7) suggest that there may be several wavelengths with the ability to discriminate between T1 and T2. These results indicate that the early detection of *Ganoderma* disease in oil-palm trees may be feasible by using first derivative data sets. Figure 8(a) and (b) shows the histograms of statistically significant wavelengths for reflectance spectra and for first derivative spectral regions.

The histograms in figure 8(a) and (b) summarize the significance of the reflectance spectra and the first derivative of the reflectance spectra data sets. These histograms identify which wavelengths can potentially discriminate between all three classes (T1, T2, and T3).

![Histograms of significance frequencies](attachment:figure8.png)

**Figure 8.** Histograms of significance frequencies: (a) histogram for reflectance spectra per wavelength; (b) histogram for first derivative of reflectance spectra per wavelength.
Table 1. The most significant wavelengths as identified by ANOVA testing.

<table>
<thead>
<tr>
<th>Data type</th>
<th>Best Bands (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reflectance Spectra</td>
<td>none</td>
</tr>
<tr>
<td>First Derivative of Reflectance Spectra</td>
<td>495, 495.5, 496, 651.5, 652, 652.5, 653, 653.5, 654, 654.5, 655, 655.5, 656, 656.5, 657, 657.5, 658, 658.5, 659, 659.5, 660, 660.5, 661, 908</td>
</tr>
</tbody>
</table>

T2, T3). The most responsive wavelengths are defined by regions with a maximum frequency value of close to 3.

From figure 8(a), we conclude that no wavelengths offer the greatest frequency of significance. The maximum frequency of significance in figure 8(a) can only explain the significant difference between T1 and T3, or between T2 and T3. In figure 8(b), 24 wavelengths have the highest frequency of significance, indicating that there may be a significant difference between all three possible pairs involving T1, T2 and T3.

The 24 most significant wavelengths (as listed in table 1) were located in the green (495, 495.5 and 496 nm), red (651.5, 652, 652.5, 653, 653.5, 654, 654.5, 655, 655.5, 656, 656.5, 657, 657.5, 658, 658.5, 659, 659.5, 660, 660.5 and 661 nm) and infrared (908 nm) spectral regions. As mentioned in Carter (1994), green and red bands are sensitive to the chlorophyll content, while the infrared region is sensitive to the collapsed structure of the green leaves. The locations of the significant wavelengths in these regions confirm that *Ganoderma* affects the chlorophyll content and the leaf structures of the oil-palm tree leaves.

However, these significant wavelengths were not reported in the results published by Fung *et al.* (2003), which only mentioned significant spectral regions in the green peak (500–570 nm) for subtropical tree species. Our study suggests that the green peak is not sensitive to the health of oil-palm trees, although it has proven significant for intra-species identifications. This result also did not concur with results obtained by Wang and Sousa (2009) that suggest that the near-infrared plateau (780–810 nm) is significant for inter-species discrimination among mangroves. They conclude that only one spectral location (908 nm) in the infrared region is significant for discriminating between the three health classes. Of the 24 most significant wavelengths, some are located in the red-edge region (650–750 nm). This result was consistent with a previous study conducted by Fung *et al.* (2003) that proposed the red-edge region as significant for intra-species discrimination.

These results suggest that future research may benefit from correlation analyses between first derivative data sets and the oil-palm leaves’ chlorophyll content and structure.

### 3.3 Spectral separability

As mentioned previously, the JM distance is used for spectral separability. From the one-way ANOVA test, the 24 most significant wavelengths were selected as inputs for the JM distance measurements. The JM distance was calculated for each pair of classes (T1 vs. T2, T2 vs. T3 and T1 vs. T3). The average JM distance was calculated on the 24 wavelength combinations to determine the separation distance of the classes from one another. The JM distances ranged from 0 to 1.4142, with higher values representing better separability for all class pairs. Therefore, the wavelengths or wavelength
Table 2. J-M distance for the first derivative of the reflectance spectra (24 bands).

<table>
<thead>
<tr>
<th>Class pair</th>
<th>Average J-M distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1 vs. T2</td>
<td>1.4142</td>
</tr>
<tr>
<td>T1 vs. T3</td>
<td>1.4142</td>
</tr>
<tr>
<td>T2 vs. T3</td>
<td>1.4142</td>
</tr>
</tbody>
</table>

Table 3. Average J-M distances for the most significant wavelengths.

<table>
<thead>
<tr>
<th>Wavelength (nm)</th>
<th>Average J-M distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>495</td>
<td>0.4122</td>
</tr>
<tr>
<td>495.5</td>
<td>0.3946</td>
</tr>
<tr>
<td>496</td>
<td>0.3804</td>
</tr>
<tr>
<td>651.5</td>
<td>0.4038</td>
</tr>
<tr>
<td>652</td>
<td>0.4162</td>
</tr>
<tr>
<td>652.5</td>
<td>0.4288</td>
</tr>
<tr>
<td>653</td>
<td>0.4423</td>
</tr>
<tr>
<td>653.5</td>
<td>0.4549</td>
</tr>
<tr>
<td>654</td>
<td>0.4680</td>
</tr>
<tr>
<td>654.5</td>
<td>0.4782</td>
</tr>
<tr>
<td>655</td>
<td>0.4862</td>
</tr>
<tr>
<td>655.5</td>
<td>0.4911</td>
</tr>
<tr>
<td>656</td>
<td>0.4914</td>
</tr>
<tr>
<td>656.5</td>
<td>0.4911</td>
</tr>
<tr>
<td>657</td>
<td>0.4877</td>
</tr>
<tr>
<td>657.5</td>
<td>0.4848</td>
</tr>
<tr>
<td>658</td>
<td>0.4798</td>
</tr>
<tr>
<td>658.5</td>
<td>0.4719</td>
</tr>
<tr>
<td>659</td>
<td>0.4643</td>
</tr>
<tr>
<td>659.5</td>
<td>0.4486</td>
</tr>
<tr>
<td>660</td>
<td>0.4350</td>
</tr>
<tr>
<td>660.5</td>
<td>0.4155</td>
</tr>
<tr>
<td>661</td>
<td>0.3849</td>
</tr>
<tr>
<td>908</td>
<td>0.3470</td>
</tr>
</tbody>
</table>

combinations that produce the highest JM distances averaged over all class pairs can be considered optimal for discrimination purposes. In addition, the JM distance of individual significant wavelengths was also calculated to determine the separability of individual wavelengths.

Table 2 shows the resulting JM distance averaged across all combined class pairs, together with the separability indexes of the 24 most significant wavelength combinations that yield the highest separability values. Table 3 presents the average JM distances for the individual wavelengths. The best separability comes from using a single wavelength located at 656 nm, but it produces an unacceptable JM value of 0.4914. On the other hand, the use of 24 wavelengths may give the best class separability.

Table 3 shows that the highest JM value for any single wavelength is at 656 nm, in the red band. However, the JM value for this wavelength does not achieve an
Table 4. Error matrix for the maximum likelihood classification involving the 24 most significant wavelengths.

<table>
<thead>
<tr>
<th></th>
<th>T1</th>
<th>T2</th>
<th>T3</th>
<th>Row total</th>
<th>User’s accuracy (%)</th>
<th>Commission error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>23</td>
<td>0</td>
<td>1</td>
<td>24</td>
<td>96</td>
<td>4</td>
</tr>
<tr>
<td>T2</td>
<td>5</td>
<td>15</td>
<td>4</td>
<td>24</td>
<td>63</td>
<td>37</td>
</tr>
<tr>
<td>T3</td>
<td>3</td>
<td>0</td>
<td>21</td>
<td>24</td>
<td>88</td>
<td>12</td>
</tr>
<tr>
<td>Column total</td>
<td>31</td>
<td>15</td>
<td>26</td>
<td>72</td>
<td>74</td>
<td>100</td>
</tr>
<tr>
<td>Producer’s accuracy (%)</td>
<td>74</td>
<td>100</td>
<td>81</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Omission error (%)</td>
<td>26</td>
<td>0</td>
<td>19</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

acceptable separation between classes. Thus this individual wavelength cannot be used for classifying any three classes. Furthermore, the JM distance does not provide any information about the classification accuracy of the most significant wavelengths from the first derivative spectra. Therefore, the spectral responses of the most significant wavelengths should instead be used for classification purposes in connection with the maximum likelihood classifier.

### 3.4 Maximum likelihood classification and accuracy assessment

All of the 24 most significant wavelengths were input into a maximum likelihood classifier. This process examined the accuracy of the classification of the three health classes using the most significant wavelengths. An error matrix was computed for these results. Omission and commission errors together with producer’s and user’s accuracies were also calculated and are listed in table 4. The commission error of T2 is larger than that of T1 and T3 due to the overlaps of the spectral responses of T2 with T1 and T3. The overlapping spectral responses did cause minor confusion within the classification engine.

We conclude that the net classification accuracy reached an acceptable value of 82% and a Kappa coefficient of 0.73 (73%), both of which suggest moderate classification accuracy. These moderate results are believed to derive from the spectral response overlaps in the data sets. This result proves that combinations of the 24 significant wavelengths offer an acceptable, moderate ability to discriminate between the three classes.

### 4. Discussion

Numerous studies have examined the ability of hyperspectral data to identify damage caused by disease and pathogens. This study reports on disease detection in oil-palm trees using hyperspectral technology, a largely unexplored research area to date. This study proves that the first derivative of the reflectance spectra improved the ability of hyperspectral data to discriminate between healthy and infected oil-palm trees. Furthermore, several wavelengths that are located in the blue, green, red and infrared spectra regions of the first-derivative data sets may help better discriminate between healthy and infected oil palms. Detection of infection in plants without foliar symptoms indicates that early detection of *Ganoderma* disease in oil-palm trees using hyperspectral data is feasible. The results in this study do agree with other published
data (Schmidt and Skidmore 2002, Naidu et al. 2009) that indicate that the visible, red and infrared regions do feature important wavelengths for disease detection in vegetation.

The significant differences that exist in the red region indicate that this disease is also related to the chlorophyll content of oil-palm leaves. Wavelengths within the 690 – 700 nm regions are particularly sensitive to decreases in leaf chlorophyll content and represent a blue shift of the red edge that frequently accompanies stress (Cibula et al. 1992, Carter and Knapp 2001). Cellular structure and water content are the main determinants in the infrared region (Kalacska et al. 2007), and our data may be an indicator of water stress suffered by the infected oil-palm leaves.

The JM distance calculated using all 24 of the most significant wavelengths indicate that the band combinations yield an average separability of 100% for T1 versus T2, T2 versus T3 and also T1 versus T3. The results confirm that the average separability is higher than the acceptable threshold of separability, which is ≥95%. The JM results are further supported by the maximum likelihood classification results, which indicate an acceptable moderate classification accuracy. The net accuracy was 82% and the Kappa statistic value was 0.72 (72%).

The findings encourage the development of new indices that can improve the separability between the three classes and that could in theory be applied for use in airborne and spaceborne hyperspectral sensors.

5. Conclusion

In this article, we examined the ability to discriminate between three classes of *Ganoderma* disease severity (healthy, mild symptoms and severe symptoms coded by T1, T2 and T3, respectively) by using hyperspectral remote-sensing data. Hyperspectral data were used to discriminate between these classes with various separability distances and variable degrees of success.

We used one-way ANOVA to analyse the hyperspectral data, and we tested both the reflectance spectra and the first derivative of the reflectance spectra. Our ANOVA results identified wavelengths that can detect differences between the classes. This study indicates that oil-palm trees that are infected by *Ganoderma* disease can be detected using certain wavelengths within the blue, green, red and infrared regions. We conclude that the first derivative of the spectral reflectance data can help identify additional wavelengths that should unveil significant differences between the classes.

The utility of the chosen wavelength combinations was inspected by calculating the separation distance between classes using the JM distance measure. Our results show that the separability distances between all class pairs achieved a maximum separability index value of 1.4142. The classification accuracy assessments of the three classes indicate that the spectral responses of the 24 most significant wavelengths are all moderate.

In summary, hyperspectral data may help discriminate between healthy and *Ganoderma*-infected oil palms. The use of the first derivative of the reflectance spectra can improve the separation between the classes. We conclude that the first derivative of the reflectance spectra is more effective than the reflectance spectra in the early detection of *Ganoderma*. Future work will include further improvement of the hyperspectral signal analysis through the development of spectral indices based on those wavelengths that are determined to be most significant.
Acknowledgements
The authors thank the Ministry of Higher Education Malaysia for providing the research grant under the Fundamental Research Grant Scheme FRGS and UPM for providing graduate scholarship in completing the research. Malaysian Palm Oil Board (MPOB) is also acknowledged for their cooperation and assistance in this research.

References


Spectral discrimination of oil palm using hyperspectral data


