A SIGNAL RESTORATION METHOD FOR THE INFRARED SPECTRAL REFLECTANCE OF SPECTORADIOMETER MEASUREMENT

Chinsu Lin, Professor and Department Chairman
Khongor Tsogt, Doctoral student
Department of Forestry and Natural Resources
National Chiayi University
Chiayi, Taiwan, ROC 60004
chinsu@mail.nchu.edu.tw
s0961475@mail.nchu.edu.tw

ABSTRACT

ASD Spectroradiometers are widely used for objective spectral measurements in many scientific fields for years. In the field use of forest research, we have a problem of signal noise in the infrared area, especially in water band which is probably caused by weak solar intensity and/or high air moisture content. Signal noise ratio (SNR) of ASD spectroradiometer is decreased and hence make fatal data applications. According to a number of our field experiments, we found that a noisy reflectance usually occurred in the area ranges from 1350 to 1410 nm and from 1800 to 2000 nm which caused a problem in exploring the spectral characteristics of tree physiology. So, this paper aimed to resolve the noise problem by empirical modeling via field experiments of tree’s spectra for further potential uses; and only the first noisy region, i.e., 1350-1410 nm (NIR-1350-1410 nm) was studied.

Conifer Red Cypress trees which grow in Alishan Mountain were selected for such experimental measurements, including on field and in lab measurements. Empirical models were developed based on the lab spectra and then applied to the field spectra with noisy reflectance. Results demonstrated that sigmoid and logistic models could depict the reflectance behavior in the region of 1350-1410 nm of the Red Cypress with an R² of 0.99. However, the models contain significant bias which is wavelength dependent and normally distributed. It was found that integrating the Gaussian model with the sigmoid/logistic model could fit the spectral signal or noise free reflectance curve very well. Based on the test of goodness of fit, it is suggested that a 4-parameter logistic signal model combined with a 3-parameter Gaussian bias model is suitable for the application of noise removal for field measurement.

Keywords: signal restoration, noise removal, empirical models, hyperspectral remote sensing, ASD spectroradiometer.

INTRODUCTION

Hyperspectral remote sensing is a young science and its advantage is to explain material and natural phenomenon by hyperspectra at every nanometer between 350-2500 nm. Hyperspectral remote sensing is still developing and most of the users who do research in this field are trying to draw necessary information and to build a spectral library of natural objects for purposes. For this case spectral data should be accurate.

The atmosphere absorbs much of the sun’s ultraviolet radiation, and water vapor absorbs most of the incoming near infrared light around 1400, 1900 and greater than 2700 nm bands (Govender, 2007). This leads to a normal decrease in signal to noise ratios (SNR). Data quality in analog and digital communications depend on signal to noise ratio and in remote sensing it is affected by atmospheric gases. Water absorption features for liquid water can be found at approximately 970, 1200, 1450 and 1950 nm (Curran, 1989). The features at 1450 and 1950 nm are most pronounced. However, at about 1400 and 1900 nm, broad absorption features also occur due to water vapor in the atmosphere. As a result, hardly any radiation is reaching the Earth’s surface and, thus, the liquid water bands at 1450 and 1950 nm cannot be used (Clevers et al., 2006).

Hyperspectral sensor developers are trying to solve this problem and offer some suggestions. Clark et.al. (2002) suggested a possible way to reduce atmospheric effects is to use an artificial light source, powerful enough to measure a large spot in an enclosure to minimize solar and atmospheric effects. However, batteries for power lights and the spectrometer, which last long enough to measure many spots, would be cumbersome to carry, especially on rough terrain. One solution to this difficulty with field measurements is to sample the surface materials and measure
their spectra in the laboratory. The U.S. Geological Survey has found that for some calibration sites, the laboratory spectra of field samples usually measure enough data of the site's spectral characteristics to enable determination of artifacts and real features in the field data. Thus, the laboratory spectral data provides a secondary check to establish the quality of the in situ field.

ASD suggest that an averaging method can be used to reduce noise in the desired spectrum because noise itself is very natural and in reality, it has no actual effect. Most researchers reduce noise through averaging spectra (10 to 25 is sufficient) both in real-time and in post-processing. Too much spectra averaging takes time, causes delay and even clear sky atmospheric conditions always change frequently. The simplest solution is to use Median values in post-processing, rather than averages. Such as those in water bands and other saturated absorption bands will be more accurate when using the median than the mean (ASD, 2002). However, spectrum averaging or using median values can improve signal to noise ratio in dry atmospheric conditions but there is still no way to get noiseless spectra at normal and high moisture content atmospheric conditions (ASD, 1999). According to our experience if atmosphere contains a high concentration of humidity (Fig. 1), the method of spectrum averaging or median values was not sufficient for noise reduction (Fig. 2).

**Figure 1.** Average humidity of the last 5 year (2004-2008) on Alishan mountain, Taiwan.

**Figure 2.** Red Cypress reflectance at the Alishan mountain, Taiwan. 30 spectrum averaging results from a single measurement in the region 350-2500 nm (*left*), and an averaged spectrum of 5 measurements (each measurement had taken 30 spectrum averaging) and median values of first short wave infrared region at 1330-1430 nm (*right*).

Some researchers do not use data sets which are located in the spectral region of atmospheric water absorption (Clevers et al., 2006; Ma and Chen, 2005; Psomas et al., 2005), although it is recognized that this area in the spectrum contains predictive information about liquid water. For this reason, it is impossible to do analysis with noisy reflectance data. So, the signal problem should be reconsidered.

Considering the spectra of natural targets, a mathematical model is suitable for describing the generalized spectral behavior of objects. Under this assumption, this paper applies linear, sigmoid, logistic models and integrates them with the Gaussian model to figure out the spectral behavior in the infrared region ($\lambda_{1350-1410}$ nm) of the Red Cypress tree canopy. And then the models are used to fix the noisy spectral data in that specific infrared region.
EMPIRICAL REGRESSION MODELS

Based on our experiments, noisy spectra of the Red Cypress always occur at the region from 1350 to 1410 nm. We therefore need noise-free extended spectra data for modeling the reflectance/signal curve in the region. 20 nm each side was extended in this study. Several models, including (i) the simple linear model, (ii) the four-parameter sigmoid model, and (iii) the four-parameter logistic model were used to fit reflectance curves. Then (iv) the three-parameter Gaussian model was used to correct bias compared with the original data for each model.

(i) Simple linear model. Simple linear function is a function with a single regressor \( x \) that has a relationship with a response \( y \) that is a straight line. Such a function can be written as:

\[ y = y_0 + ax \]  
(Eq. 1)

(called slope-intercept form), where \( y_0 \) and \( a \) are real constants and \( x \) is a real variable. The constant \( a \) is often called the slope while \( y_0 \) is the y-intercept, which gives the point of intersection between the graph of the function and the y-axis. Changing \( a \) makes the line steeper or shallower, while changing \( y_0 \) moves the line up or down (Montgomery, 2001).

(ii) Four parameter sigmoid model. A sigmoid function is a curve having an "S" shape. In general, a sigmoid function is real-valued and differentiable, having either a non-negative or non-positive first derivative and exactly one inflection point. There are also two asymptotes, \( x \to \pm \infty \).

\[ y = y_0 + \frac{a}{1 + e^{\frac{x-x_0}{b}}} \]  
(Eq. 2)

The constant \( a \) is often called the slope while \( y_0 \) is the y-intercept, which gives the point of intersection between the graph of the function and the y-axis. When the constant of \( b \) is positive the curve increases from left to right. Conversely, when the constant of \( b \) is negative the curve decreases from left to right. The parameter \( x_0 \) sets the point of inflection of the curve (Mitchell, 1997).

(iii) Four parameter logistic model. Besides the logistic function, sigmoid functions include the ordinary arc-tangent, the hyperbolic tangent, and the error function. This function has been detailed extensively (Mitchell, 1997) and is employed in a wide variety of scientific disciplines. Standard curves commonly display a pronounced sigmoidal shape when plotted on an \( x \) and \( y \) axis. The logistic function adopts the same general shape and is a reasonable relationship to use in modeling standard curves. It is described as:

\[ y = \min + \frac{\text{max} - \text{min}}{1 + \left( \frac{x}{\text{Hillslope}} \right)^{\text{Hillslope}}} \]  
(Eq. 3)

Parameters \( \text{min} \) and \( \text{max} \) represent the upper and lower plateaus of the curve. The \( EC_{50} \) is the inflection point, which is the \( x \) value when \( y \) is half-way between the \( \text{min} \) and \( \text{max} \) plateaus and associated with the point of symmetry of the sigmoid. \( \text{Hillslope} \) is a curvature parameter and is related to the slope of the curve.

(iv) Three parameter Gaussian model. Gaussian functions are widely used in statistics where they describe the normal distribution (Weibull++7, 2005). Gaussian function is a function of the form:

\[ y = a e^{-0.5 \left( \frac{x-x_0}{b} \right)^2} \]  
(Eq. 4)

where the real constants \( a > 0 \), \( x_0 \), \( b > 0 \), and \( e \approx 2.718281828 \) (Euler's number).

The graph of a Gaussian curve is a characteristic symmetric "bell shape curve" that quickly falls off towards plus or minus infinity. The parameter \( a \) is the height of the curve's peak, \( x_0 \) is the position of the center of the peak, and \( b \) controls the width of the "bell".
MATERIALS AND METHODS

The study site is a plantation of Red Cypress trees (Chamaecyparis formosensis Obtusa) which are located on Alishan Mountain in central Taiwan at N 23°29'56", E 120°49'08" with an average elevation of 2267 m. Red Cypress is an endemic species and is one of the most valuable wood resources in Taiwan. On October 9th, 2008, a field survey with an ASD FieldSpec Pro FR spectroradiometer was carried out. The experimental plot had 6 trees and a watchtower (height 6.5 m) is located between these trees in order to measure reflectance. The spectroradiometer was deployed using a fiber optic cable with a 25° field of view. Measurement height above the crown was about 1.5-2 m. As a result, the field of view at the crown level was circular with a radius ranging from 0.66 – 0.89 m. Spectral data was collected under cloud-free conditions between 09:20 and 10:00 hrs standard local time. Target measurements were collected immediately after the reference measurements. 5 replications of spectroradiometer measurements per tree were taken, whereby each measurement represents the average of 30 readings at the same crown and detailed field protocol was adhered to. The sampling interval was 1 nm. Calibration was done by using a Spectralon white reference panel every 20 minutes to account for the changing state of the visible near infrared detector, removing the effects of dark current. The spectral reflectance of the trees was calculated using reflected target radiance divided by the irradiance of the white panel. A total of 30 spectral measurements were observed but due to tower shadow effect, 10 of them were excluded from this study. After the field survey, 30 samples of Red Cypress trees were taken for spectral measurement in laboratory.

The research was conducted to calculate the normally accurate reflectance for fixing the noisy spectral data at NIR_{1350-1410 nm} region of targets. The procedure applied to meet this end was firstly to build empirical signal models using noise-free laboratory measurements at NIR_{1330-1430 nm} region (Fig. 3). Secondly, to derive the models’ estimation error and to build relative bias models for error correction. And finally to integrate both signal and bias models in order to restore the spectral reflectance of the field data.

An analysis of variance (ANOVA) test and Student’s t-test were applied to examine the significance of regression models and their coefficients ($\beta$). Variance inflation factor (VIF) is an indicator of multicollinearity in regression model. It measures how much the variance of a regression coefficient (square of the standard error) is increased because of multicollinearity. Larger VIF values will make the empirical model unstable and give poor prediction/estimation. A common rule of thumb is that if VIF($\beta$) > 5 then multicollinearity is high, and a value of 10 has been proposed as a cut off value (Longnecker and Ott, 2004).

RESULTS AND DISCUSSIONS

Figure 4 depicts the reflectance curves of both the laboratory and field cases. The laboratory spectra was relatively more stable than the field spectra. Median values for each wavelength in the NIR_{1330-1430 nm} are visible in the middle portion of the spectral reflectance belt in the laboratory cases. The median and mean value are very similar in the left graph of figure 4. While in the field cases shown in the right graph of figure 4, the median value (blue line) showed a more fluctuating curve and did not match with the mean curve (dashed line). Field spectra of Red Cypress are significantly affected by the light environment that is why the spectra belt in the right graph of figure 4 shows an irregular pattern.
Empirical Signal and Bias Models of Red Cypress

The fitted empirical signal models that accompany the regression coefficients and VIFs are shown in Table 1. Those models were statistically significant with an ability of 99% reflectance explanation ($R^2=0.99$). Every regression coefficient was significant at the probability level $p<0.0001$, but VIF of the coefficients varied dramatically among these models. The simple linear signal model whose intercept and slope estimates’ VIF equal to 1164.58 was concluded to be the poorest one among those developed models. The four parameter sigmoid model and the four parameter logistic model were the better ones. It is obvious that the coefficients of the logistic signal model have smaller VIFs but it seems that the logistic model is better than the sigmoid because of whose coefficients’ VIF are almost less than or close to the multicollinearity threshold values. Particularly the VIF of the slope estimate of the logistic model which is 7.52. This is less than 24.12 which is the VIF of slope of the sigmoid model.

Table 1 showed the goodness of fit indicators for the prediction errors of the empirical signal models. The weakness of the linear signal model can be diagnosed by the $R^2 (=0.72)$ of its Gaussian bias model. The prediction errors of the empirical sigmoid and the logistic models are very similar and could well be explained by a 3-parameter Gaussian model whose $R^2$ is 0.99. The VIF of every regression coefficient is very small. According to Table 1 and Table 2, it is suggested that the reflectance of Red Cypress could well be predicted by integrating both the four-parameter logistic/sigmoid model and the three-parameter Gaussian model; the combination of Logistic-Gaussian models is the most suitable method.

### Table 1. Signal models comparison for laboratory measurements ($R^2=0.99$ and $p<0.0001$)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Coefficient</th>
<th>VIF</th>
<th>Parameter</th>
<th>Coefficient</th>
<th>VIF</th>
<th>Parameter</th>
<th>Coefficient</th>
<th>VIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y_0$</td>
<td>3.6092</td>
<td>1164.58</td>
<td>$x_0$</td>
<td>1380.048</td>
<td>13.16</td>
<td>min</td>
<td>0.1252</td>
<td>17.02</td>
</tr>
<tr>
<td>$a$</td>
<td>-0.0024</td>
<td>1164.58</td>
<td>$y_0$</td>
<td>0.1259</td>
<td>48.70</td>
<td>max</td>
<td>0.3663</td>
<td>1.27</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$a$</td>
<td>0.2410</td>
<td>59.65</td>
<td>EC50</td>
<td>1380.082</td>
<td>10.27</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$b$</td>
<td>-16.956</td>
<td>24.12</td>
<td>Hillslope</td>
<td>81.3998</td>
<td>7.52</td>
</tr>
</tbody>
</table>

### Table 2. Bias Gaussian models comparison for bias correction ($p<0.0001$)

<table>
<thead>
<tr>
<th>Model’s Parameter</th>
<th>(a) Simple linear model ($R^2=0.72$)</th>
<th>(b) Four parameter sigmoid model ($R^2=0.99$)</th>
<th>(c) Four parameter logistic model ($R^2=0.99$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient</td>
<td>VIF</td>
<td>Coefficient</td>
</tr>
<tr>
<td>$a$</td>
<td>0.0539</td>
<td>1.5</td>
<td>0.0532</td>
</tr>
<tr>
<td>$b$</td>
<td>22.510</td>
<td>1.5</td>
<td>12.865</td>
</tr>
<tr>
<td>$x_0$</td>
<td>1370.75</td>
<td>1</td>
<td>1379.44</td>
</tr>
</tbody>
</table>
Figure 5. The simple linear-Gaussian method. Original reflectance and the estimates of the empirical model (top-left), prediction error of the fitted linear model (top-right), Gaussian bias model and its prediction band (bottom-left), and the restored final reflectance with respect to the original reflectance (bottom-left).

The blue curve in Figure 5 (top-left) shows the average reflectance of the original laboratory samples. Using the reflectance at two ends (\( \rho_{1330-1350\text{nm}} \) and \( \rho_{1410-1430\text{nm}} \)) of the curve, a green line was determined by the simple linear model. It demonstrated an N-shape distribution of the prediction bias (Figure 5, top-right). The prediction error of the fitted linear signal model was distributed around the zero value. A constant adjustment factor with a value of 0.02 was incorporated with the signal error making it available for the Gaussian bias modeling. The fitted 3-parameter Gaussian model showed a very wide distributed prediction band for such wavelength dependent bias (bottom-left); it indicates that the bias estimation deviates a little from the bias produced by the linear signal model. From the bottom-right graph of Figure 5, the deviation occurred in the range 1370-1430 nm. Although it seems that a three-order polynomial model would fit the bias distribution, it is not used based on the general application and moreover the VIF of model coefficients were extremely large and not accepted.

Figure 6 and Figure 7 demonstrate the original laboratory spectra and the empirical model estimated spectra curves (top), bias and the prediction band of the bias model (middle) of the sigmoid model and that of the logistic model. The restored signal and its deviation from the original reflectance signal for those two empirical models are shown in the bottom of Figure 6 and 7. Comparing the paired graphs of Figure 6 and 7, indicates that these two empirical models have similar efficiencies in the signal restoration of the water absorption region of Red Cypress if integrated with the 3-parameter Gaussian bias model. It is suggested that Figures 5, 6 and 7 have shown that both logistic-Gaussian and sigmoid-Gaussian modeling methods are visually better than the linear-Gaussian method.
Application of Signal Restoration Methods

Experimental results of the laboratory spectra have shown that the logic of integrating mathematical models is a useful algorithm for the effective modeling of tree’s reflectance curve in the water absorption spectral region. This section introduced the algorithm into the field data of Red Cypress.

Table 3 lists the coefficients and their VIFs of those empirical signal models of the field data set. The results were almost identical to the laboratory case listed in Table 1. Coefficients for linear, sigmoid, and logistic signal models of the field case were very close to the laboratory case; and VIFs of every coefficient are also in the similar situation. Those Gaussian bias models derived respectively from linear, sigmoid, and logistic models of the laboratory data set were directly applied to correct the corresponding reflectance of each signal model.
Results of the signal restoration were demonstrated in Figure 8. Obviously, the integration of the logistic signal model and the Gaussian bias model showed the best correction effect with respect to the goodness of fit. Restored reflectance in the spectral region ($\lambda_{1330-1430}$nm) exhibited a curve tendency which follows the curve trend of the original observed reflectance. While the combination of the linear signal model with the Gaussian bias model or the sigmoid signal model with the Gaussian bias model showed some deviation from the restored reflectance and the observed reflectance. Specifically, the restored reflectance was over-estimated by the linear model and underestimated by the sigmoid model.

We know from the section on laboratory reflectance modeling that the sigmoid and the logistic models were visually similar to each other; but the similarity could not be kept in the field reflectance. This might be due to a potential deviation problem of the sigmoid model; and it is perhaps not possible to determine accurately the bias of signal reflectance curves. Since the Gaussian model determined bias adjustment for the signal of empirical logistic model could work very well to match the original reflectance; it is therefore suggested that the four parameter logistic model integrated with the three parameter Gaussian model is the best method to fix noisy spectra of the near infrared region from 1350 to 1410 nm.

| Table 3. Signal Model comparison for the field data set ($R^2=0.99$ and $p<0.0001$) |
|------------------------------------------|------------------------------------------|------------------------------------------|
| Parameter | Coefficient | VIF  | Parameter | Coefficient | VIF  | Parameter | Coefficient | VIF  |
| $y_0$     | 3.5644      | 1164.58 | $x_0$     | 1375.251    | 12.18 | $\text{min}$ | 0.0671      | 13.21  |
| $a$       | -0.0024     | 1164.58 | $y_0$     | 0.0674      | 35.98 | $\text{max}$ | 0.3053      | 1.26   |
|           |             |        | $a$       | 0.2379      | 50.13 | $E_{C_{50}}$ | 1375.229    | 10.52  |
|           |             |        | $b$       | -15.996     | 23.17 | $\text{Hillslope}$ | 86.3627 | 9.30 |

**Figure 8.** Differences between the field average reflectance and the linear model (*upper left*), the four parameter sigmoid model (*upper right*), and the four parameter logistic model (*left*) with the addition of Gaussian bias model reflectance.

**CONCLUSION**

This paper applied integrated mathematical models, such as linear, sigmoid, logistic and Gaussian models, to explain the spectral behavior of the NIR$_{1350-1410}$nm area of a Red Cypress tree and then used them to fix the noisy spectral data in that specific infrared region. Results showed that both the four-parameter sigmoid model and the four-parameter logistic model can match the curve of the NIR$_{1350-1410}$nm of reflectance and more than $R^2=0.99$ of the...
spectral values could be explained using only these empirical models. Considering the variance inflation factor (VIF) of the regression coefficients, it is recommended that the logistic model is better than the sigmoid model, since the parametric coefficients of the logistic model were statistically tested and found to be acceptable. The coefficients of the logistic model almost satisfied the boundaries of collinearity criteria while the derived sigmoid model did not. Estimation biases of the logistic signal model are wavelength dependent. Using a three-parameter Gaussian model could modify such biases effectively.

Primary results from this study demonstrate that the four parameter logistic model provides good opportunities for describing signal reflectance at NIR 1350-1410 nm region of the Red Cypress. In addition, this study showed that the Gaussian bias model of laboratory spectral data can improve the signal model fitness. Briefly, The proposed method, integrates the 4-parameter logistic signal model and the 3-parameter Gaussian bias model. This will bring benefits to foresters in the field conducting spectral experiments where air moisture is high and/or the incidence of solar intensity is low. Use of the proposed algorithm, can improve the quality of spectroradiometer field data and can be expected to satisfy the ends of hyperspectral remote sensing applications.

REFERENCES