ESTIMATION OF LEAF AREA INDEX AND CHLOROPHYLL FOR A MEDITERRANEAN GRASSLAND USING HYPERSONTRAL DATA

R. Darvishzadeh a,b*, A. Skidmore a, M. Schlerf a, C. Atzberger c, F. Corsi d, M. Cho e

a International Institute for Geo-information Science and Earth Observation (ITC), Hengeloestraat 99, P.O. Box 6, 7500 AA Enschede, The Netherlands - (darvish, skidmore, schlerf)@itc.nl
b Department of Remote Sensing and GIS, Faculty of Earth Sciences, Shahid Beheshti University, Tehran, Iran – darvishzade@yahoo.com
c Joint Research Centre of the European Commission, TP 266, Via Enrico Fermi 1, 21027 Ispra (VA), Italy - clement.atzberger@jrc.it
d 712 Locksley Rd, Yorktown Heights, NY, 10598, USA - fabio@st-fc.org
e Ecological Remote Sensing, Natural Resources and the Environment, Earth Observation Group, Building 33 CSIR, Pretoria, SA - Mcho@csir.co.za

Commission VII, WG VII/3

KEY WORDS: Hyperspectral, Vegetation, Estimation, Leaf Area Index, Chlorophyll

ABSTRACT:

The study shows that leaf area index (LAI) and canopy chlorophyll content can be mapped in a heterogeneous Mediterranean grassland from canopy spectral reflectance measurements. Canopy spectral measurements were made in the field using a GER 3700 spectroradiometer, along with concomitant in situ measurements of LAI and chlorophyll content. We tested the utility of univariate techniques, involving narrow band vegetation indices and the red edge inflection point, as well as multivariate calibration techniques, such as partial least squares regression. Among the various investigated models, canopy chlorophyll content was estimated with the highest accuracy ($R^2_{CV} = 0.74$, relative RMSE$_C = 0.35$) and LAI was estimated with intermediate accuracy ($R^2_{CV} = 0.67$). Compared with narrow band indices and red edge inflection point, partial least squares regression generally improved the estimation accuracies. The results of the study highlight the significance of using multivariate techniques such as partial least squares regression rather than univariate methods such as vegetation indices for providing enhanced estimates of heterogeneous grass canopy characteristics. To date, partial least squares regression has seldom been applied for studying heterogeneous grassland canopies. However, it can provide a useful exploratory and predictive tool for mapping and monitoring heterogeneous grasslands.

1. INTRODUCTION

Owing to its fast, non-destructive and relatively cheap characterization of land surfaces, remote sensing has been recognized as a reliable method for estimating various biophysical and biochemical vegetation variables (Curran et al., 2001; Hansen and Schjoerring, 2003; Weiss and Baret, 1999). Hyperspectral remote sensing with narrow and continuous spectral bands that provide an almost continuous spectrum is considered more sensitive to specific vegetation variables such as leaf area index (LAI) (Hansen and Schjoerring, 2003). Because of the role of green leaves in controlling many biological and physical processes of plant canopies, LAI (the total one-sided leaf area per ground surface area) is a key structural characteristic of vegetation and thus widely used as an indicator of vegetation status.

LAI has been estimated in numerous studies by using remote sensing in either statistical approaches or physically based (canopy reflectance) models. Many of the previous studies, however, are based on simulated data (Atzberger, 2004; Broge and Leblanc, 2001; Haboudane et al., 2004), on agricultural crops (Atzberger, 1995; Atzberger, 1997; Baret et al., 1987; Broge and Mortensen, 2002; Jacquemoud et al., 2000; Walter-Shea et al., 1997; Weiss et al., 2001) or on forest (Chen et al., 1997; Fang et al., 2003; Kalacska et al., 2004; Running et al., 1986; Schlerf and Atzberger, 2006; White et al., 1997), where single species was investigated. Therefore, investigation is required to assess the capability of remote sensing models when it comes to natural heterogeneous canopies with a combination of different plant species in varying proportions. Mediterranean grasslands are characterized by highly heterogeneous canopies, and present a challenge for remote sensing applications because the reflectance is often a mixture of different surface materials (Fisher, 1997; Roder et al., 2007).

The aim of this study was to examine the utility of hyperspectral remote sensing in predicting canopy characteristics such as LAI and canopy chlorophyll content in a heterogeneous Mediterranean grassland by means of different univariate and multivariate methods. We compared narrow band vegetation indices, including red edge inflection point (REIP), with partial least squares regression. The suitability of these different methods will be analyzed in terms of their prediction accuracy. Naturally, the significance of the results is valid only for Mediterranean grasslands and the biophysical variables considered. The study is based on canopy spectral reflectance measured in a heterogeneous grassland during a field campaign in the summer of 2005 in Majella National Park, Italy.
2. METHODS

2.1 Study area and sampling

The study site is located in Majella National Park, Italy (latitude 41°52’ to 42°14’ N, longitude 13°14’ to 13°50’ E). The park covers an area of 74,095 ha and extends into the southern part of Abruzzo, at a distance of 40 km from the Adriatic Sea. The region is situated in the massifs of the Apennines. The park is characterized by several mountain peaks, the highest being Mount Amaro (2794 m). Coordinates (x y) were randomly generated in a grassland stratum to select plots. A total of 45 plots (30 m x 30 m) were generated and a GPS (Global Positioning System) was used to locate them in the field. To increase the number of samples in the time available, four to five randomly selected subplots were clustered within each plot. This resulted in a total of 191 subplots being sampled. The 1 m x 1 m subplots differed in species composition and relative abundance while the within-subplot variability was small.

2.2 Canopy spectral measurements

Fifteen replicates of canopy spectral measurements were taken from each subplot, using a GER 3700 spectroradiometer (Geophysical and Environmental Research Corporation, Buffalo, New York).

The fiber optic, with a field view of 25°, was handheld approximately 1 m above the ground at nadir position. The ground area observed by the sensor of GER had a diameter of 45 cm and was large enough to cover the center of the subplots without being influenced by the surroundings. The 15 replicate spectral measurements taken from each subplot enabled to suppress much of the measurement noise by averaging the replicate measurements. Prior to each reflectance measurement, the radiance of a white standard panel coated with BaSO4 and of known reflectivity was recorded for normalization of the target measurements. The fieldwork was conducted between June 15 and July 15 in 2005. To minimize atmospheric perturbations and BRDF effects, spectral measurements were made on clear sunny days between 11:30 a.m. and 2:00 p.m.

2.3 LAI measurements

In each subplot, LAI was non-destructively measured using a widely used optical instrument, the Plant Canopy Analyzer LAI-2000 (LICOR Inc., Lincoln, NE, USA). A detailed description of this instrument is given by LI-COR (1992) and Welles and Norman (1991). In this study, measurements were taken either under clear skies with low solar elevation (i.e., within the two hours following sunrise or preceding sunset) or under overcast conditions. The LAI measurements were taken on the same day that the canopy spectral measurements were made. To prevent direct sunlight on the sensor of LAI-2000, samples of below- and above-canopy radiation were made in the direction facing away from the sun (i.e., with the sun behind the operator), using a view restrictor of 45°. For each subplot, reference samples of above-canopy radiation were determined by measuring incoming radiation above the grass subplot (in an open area). Next, five below-canopy samples were collected and used to calculate the average LAI (Table 1).

2.4 Chlorophyll measurements

A SPAD-502 Leaf Chlorophyll Meter (Minolta, Inc.) was used to assess the leaf chlorophyll content (LCC) in each 1 m x 1 m subplot. A total of 30 leaves representing the dominant species were randomly selected in each subplot, and their SPAD readings were recorded. From the 30 individual SPAD measurements, the average was calculated (Table 1). These averaged SPAD readings were converted into leaf chlorophyll content (units: µg cm⁻²) by means of an empirical calibration function provided by Markwell et al. (1995). The total canopy chlorophyll content (CCC; units: g m⁻²) for each subplot was obtained by multiplying the leaf chlorophyll content by the corresponding LAI.

<table>
<thead>
<tr>
<th>Measured variables</th>
<th>Min</th>
<th>Mean</th>
<th>Max</th>
<th>StDev</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>LAI (m² m⁻²)</td>
<td>0.1</td>
<td>2.76</td>
<td>7.34</td>
<td>1.50</td>
<td>6.95</td>
</tr>
<tr>
<td>CCC (g m⁻²)</td>
<td>0.39</td>
<td>0.87</td>
<td>2.7</td>
<td>0.55</td>
<td>2.56</td>
</tr>
</tbody>
</table>

Table 1. Summary statistics of the measured biophysical and biochemical variables of grassland sample subplots (n=191); CCC is the canopy chlorophyll content.

2.5 Data analysis

We selected the normalized difference vegetation index (NDVI) (Rouse et al., 1974) as a representative of ratio indices, and the second soil-adjusted vegetation index (SAVI2) (Major et al., 1990) as a representative of soil-based indices, for the analysis in this study. The narrow band NDVI and SAVI2 indices were systematically calculated for all possible (584 x 584 = 341,056) band combinations between 400 nm and 2400 nm. The soil line parameters were calculated from soil spectral measurement of bare soils which were acquired from few subplots with no vegetation. We assumed that the measured soil optical properties were representative for the study area. Consequently, the soil line parameters were considered constant for all 191 subplots.

For this study, we used two methods to calculate the red edge inflection point (REIP). The linear interpolation method (Guyot and Baret, 1988) assumes that the spectral reflectance at the red edge can be simplified to a straight line centered around a midpoint between (i) the reflectance in the NIR shoulder at about 780 nm, and (ii) the reflectance minimum of the chlorophyll absorption feature at about 670 nm. First, the reflectance value is estimated at the inflection point. Then, a linear interpolation procedure for the measurements at 700 nm and 740 nm is applied to estimate the wavelength corresponding to the estimated reflectance value at the inflection point:

\[ R_{red-edge} = \frac{(R_{700} - R_{780})}{2} \]  

(1)

\[ REIP_{estimated} = 700 + 40 \left( \frac{R_{red-edge} - R_{725}}{R_{740} - R_{725}} \right) \]  

(2)

where the constants 700 and 40 result from interpolation between the 700 nm to 740 nm intervals, and \( R_{725} \), \( R_{740} \), \( R_{725} \) and \( R_{740} \) are, respectively, the reflectance values at 670 nm, 700 nm, 740 nm and 780 nm.

The linear extrapolation method (LEM) (Cho and Skidmore, 2006) is based on the linear extrapolation of two straight lines (Eqs. 3 and 4) through two points on the far-red (680 nm to 700 nm) and two points on the NIR (725 nm to 760 nm) flanks of the first derivative reflectance spectrum (D) of the red edge region. The REIP is then defined by the wavelength value at the intersection of the straight lines (Eq. 5).
Far-red line: \( D = m_1 \lambda + c_1 \) \hspace{1cm} (3)
NIR line: \( D = m_2 \lambda + c_2 \) \hspace{1cm} (4)

where \( m \) and \( c \) represent the slope and intercept of the straight lines, respectively. At the intersection, the two lines have equal wavelengths and \( D \) values. Therefore, the REIP, which is the wavelength at the intersection, is given by:

\[
RIEP = \frac{-(c_1 - c_2)}{(m_1 - m_2)}
\]

Partial least squares regression (PLSR) is a technique that reduces the large number of measured collinear spectral variables to a few non-correlated latent variables or factors while maximizing co-variability to the variable(s) of interest (Atzberger et al., 2003; Cho et al., 2007; Geladi and Kowalski, 1986; Hansen and Schjoerring, 2003). The latent variables represent the relevant information present in the measured reflectance spectra and are used to predict the dependent variables (here, biophysical and biochemical grass characteristics). As with other linear calibration methods, the aim is to build a linear model:

\[
Y = X\beta + \varepsilon
\]

where \( Y \) is the mean-centred vector of the response variable (grass characteristics), \( X \) is the mean-centred matrix of the predictor (spectral reflectance), \( \beta \) is the matrix of coefficients, and \( \varepsilon \) is the matrix of residuals.

The optimum number of factors was estimated by leave-one-out cross-validation. A common way of using cross-validation for this estimation is to select the number of factors that minimizes the RMSE (Geladi and Kowalski, 1986). To prevent collinearity and to preserve model parsimony, the condition for adding an extra factor to the model was that it had to reduce the root mean square error of cross-validation (RMSECV) by >2% (Cho et al., 2007; Kooistra et al., 2004). In addition, coefficients of determination (\( R^2 \)) between measured and predicted values in the cross-validation were used to evaluate the relationships found. The PLSR analysis was performed using the TOMCAT toolbox 1.01 within MATLAB (Daszykowski et al., 2007).

**3. RESULTS**

**3.1 Hyperspectral vegetation indices**

NDVI and SAVI2 narrow band vegetation indices were calculated from the measured canopy reflectance spectra, using all possible two-band combinations. The coefficients of determination (\( R^2 \)) between these narrow band vegetation indices and the grass canopy characteristics were computed. An illustration of these results is shown for LAI in the 2-D correlation plot in Figure 1. The meeting point of each pair of wavelengths in a 2-D plot corresponds to the \( R^2 \) value of LAI and the vegetation index calculated from the reflectance values in those two wavelengths. Based on the \( R^2 \) values in the 2-D correlation plots, band combinations that formed the best indices were determined for LAI and canopy chlorophyll content. The best performing indices and the band positions are tabulated in Table 2.

It can be observed from Table 2 that narrow band SAVI2 had somewhat higher correlations than narrow band NDVI with the studied variables. However, the coefficients of determination between the grass characteristics and the indices were relatively low. Studying regions where \( R^2 \geq 0.6 \) for LAI and canopy chlorophyll content (CCC) revealed that LAI had a strong influence on the selection of suitable bands for estimating canopy chlorophyll content. The similarity in the observed patterns is obviously due to the high correlation between the two variables (not shown).

![Figure 1. 2-D correlation plots illustrating the coefficient of determination (\( R^2 \)) between narrow band SAVI2 and LAI.](image)

<table>
<thead>
<tr>
<th>Variables</th>
<th>Narrow band VI</th>
<th>( \lambda [nm] )</th>
<th>( R^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>LAI</td>
<td>NDVI</td>
<td>1105/1229</td>
<td>0.61</td>
</tr>
<tr>
<td></td>
<td>SAVI2</td>
<td>1998/1402</td>
<td>0.64</td>
</tr>
<tr>
<td>CCC</td>
<td>NDVI</td>
<td>1141/1150</td>
<td>0.68</td>
</tr>
<tr>
<td></td>
<td>SAVI2</td>
<td>1211/1086</td>
<td>0.69</td>
</tr>
</tbody>
</table>

Table 2. Band positions and \( R^2 \) values between the best narrow band NDVI and SAVI2 (derived from 2-D correlation plots of different data sets) and grass variables.

For the best performing narrow band index, cross-validated \( R^2 \) and relative RMSE (\( \text{RMSE}_{\text{CV}} = \text{RMSE}/\text{mean} \)) were computed from linear regression models (Table 3). As can be observed from this table, compared with narrow band NDVI, narrow band SAVI2 gave slightly higher \( R^2 \) and lower RMSE values for LAI and canopy chlorophyll content. The better performance of SAVI2 compared with NDVI is probably due to the fact that SAVI2 is less sensitive to external factors such as soil background effects.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Narrow band VI</th>
<th>( R^2_{\text{CV}} )</th>
<th>( \text{RMSE}_{\text{CV}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>LAI</td>
<td>NDVI</td>
<td>0.60</td>
<td>0.34</td>
</tr>
<tr>
<td></td>
<td>SAVI2</td>
<td>0.63</td>
<td>0.33</td>
</tr>
<tr>
<td>CCC</td>
<td>NDVI</td>
<td>0.67</td>
<td>0.36</td>
</tr>
<tr>
<td></td>
<td>SAVI2</td>
<td>0.68</td>
<td>0.35</td>
</tr>
</tbody>
</table>

Table 3. Performance of narrow band vegetation indices for predicting grass variables in Majella National Park, Italy.
occur for canopy chlorophyll content greater than 2 (g m⁻²) and band NDVI. From the figure, it seems that saturation starts to measured LAI and canopy chlorophyll content using narrow Figure 2 shows the relationships between the estimated and variables. Among the studied variables, estimation of canopy chlorophyll content again yielded the highest R² values and the lowest relative RMSE of the grass variables obtained from the methods. As can be observed from the results reported in Table 2, methods. The REIP method was calculated using two methods. As can be observed from the results reported in Table 4, the relationships between measured and estimated grass variables were not reliable using any of the methods. The R² and relative RMSE of the grass variables obtained from the three methods were relatively similar.

Among the studied variables, estimation of canopy chlorophyll content again yielded the highest R² values and the lowest relative RMSE. Compared with regression models developed using the optimum narrow band indices, the REIP methods produced somewhat lower accuracies.

confirmed previous studies by researchers who suggested a strong contribution by SWIR bands to the strength of relationships between spectral reflectance and LAI (Cohen and Goward, 2004; Darvishzadeh et al., 2008; Lee et al., 2004; Nemani et al., 1993; Schlerf et al., 2005). Compared with the narrow band NDVI, the narrow band SAVI2 gave somewhat higher R² and lower relative RMSE values for LAI. This result is in agreement with that of Broge and Leblanc (2001), who used simulated data and found SAVI2 to be the best vegetation index for LAI estimation. Moreover, the narrow band SAVI2 performed relatively well for canopy chlorophyll content. This is due to the major influence of LAI in canopy chlorophyll content and also to the fact that SAVI2 is relatively insensitive to external factors such as soil background effects.

Although red edge has proved to respond more linearly to LAI and chlorophyll when compared with the classical NDVI, which often suffers from saturation problems (Danson and Plummer,

### Table 4. Performance of red edge inflection point calculated using different methods for predicting grass variables in Majella National Park, Italy.

<table>
<thead>
<tr>
<th>REIP method</th>
<th>R² cv</th>
<th>RRMSE cv</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear interpolation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LAI</td>
<td>0.49</td>
<td>0.39</td>
</tr>
<tr>
<td>CCC</td>
<td>0.56</td>
<td>0.41</td>
</tr>
<tr>
<td>Linear extrapolation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LAI</td>
<td>0.51</td>
<td>0.38</td>
</tr>
<tr>
<td>CCC</td>
<td>0.57</td>
<td>0.41</td>
</tr>
</tbody>
</table>

3.3 Partial least squares regression

The relationships between grass variables and reflectance spectra were modeled using PLSR. Cross-validated results using the entire reflectance spectra as inputs are shown in Figure 3. The optimal number of PLSR factors preventing over-fitting was selected in two ways: (i) through visual inspection of cross-validated RMSE versus the number of factors plots (not shown), and (ii) by setting the condition that adding an extra factor must reduce the RMSE (RMSE cv) by >2%. The number of factors in the final model were 4 for LAI and 5 for canopy chlorophyll content models. Compared with other methods, PLSR using entire reflectance spectra increased all R² values (R² = 0.69, 0.74 for LAI and canopy chlorophyll content, respectively) and decreased the Relative RMSE values (RRMSE = 0.32, 0.34 for LAI and canopy chlorophyll content, respectively).

4. DISCUSSION

The field experiment led to a large number of sample subplots (191) with high variations in LAI. The canopy integrated chlorophyll content (LAI x leaf chlorophyll content) strongly reflects the variability of LAI and (to a lesser extent) leaf chlorophyll content, expressed by the high inter-correlation between LAI and canopy chlorophyll content (not shown). Among the grass characteristics studied, canopy chlorophyll content was most accurately estimated by nearly all of the applied methods. The canopy chlorophyll content contains both the structure and chlorophyll information of vegetation and can be accurately estimated by canopy spectral reflectance.

The relationship between measured and estimated LAI was better explained by multivariate calibration methods (PLSR) than by univariate methods such as narrow band vegetation indices and REIP. This is because a two-wavelength index utilizes only a limited amount of the total spectral information available in hyperspectral data (Lee et al., 2004).

The bands selected as the best combination of the vegetation indices for LAI were found in the NIR to SWIR regions. This
1995), in our study wavelengths within the red edge region were almost absent.

The PLSR model appears to be a powerful alternative to univariate statistical methods (Darvishzadeh et al., 2008). Compared to the other investigated methods, it achieved relatively better results. It seems that important information will be lost by selecting only two wavelengths for narrow band vegetation indices.

![Figure 3](image_url)

Figure 3. Cross-validated prediction of grass variables in Majella National Park, Italy, using the entire reflectance spectra in partial least squares regression models. Left: estimated LAI versus measured LAI; right: for canopy chlorophyll content.

Estimation of biochemical and biophysical characteristics of heterogonous grassland with mixtures of different grass species is challenging in remote sensing (Roder et al., 2007), as the measured signal correspond to different grass species. In our study, an indicator of this was the observed high variations in the SPAD readings within a given subplot (not shown). Nevertheless, by using hyperspectral remote sensing with a large number of narrow spectral bands and powerful multivariate regression techniques, the biophysical grass characteristics could be retrieved with acceptable accuracy.

5. CONCLUSION

The most important conclusions that can be drawn from this study are as follows:

- Compared with LAI, canopy chlorophyll content was estimated with higher accuracy in all models.
- LAI was best estimated by partial least square regression which utilize more than two wavelengths from the entire spectral region (400 nm to 2500 nm) to estimate the variable of interest.
- SAVI2 is a potentially useful vegetation index for extracting canopy variables such as LAI. However, the selection of appropriate wavelengths and bandwidths is important.
- Partial least squares regression provided the most useful explorative tool for unraveling the relationship between canopy spectral reflectance and grass characteristics at canopy scale.

In summary, multivariate calibration methods, which until now have only been used in a few cases concerning the remote sensing of grasslands, can enhance estimates of different grass variables, and thus present new prospects for mapping and monitoring heterogeneous grass canopies from air- and space-borne platforms.

ACKNOWLEDGEMENTS

We would like to acknowledge the assistance of the park management of Majella National Park, Italy, and in particular of Dr. Teodoro Andrisano. We extend our gratitude to Dr. Istiak Sobhan for his assistance during the field campaign. Special thanks go to Dr. Michal Daszykowski for his assistance in applying the TOMCAT toolbox and for his valuable comments.

REFERENCES


