Inversion of a radiative transfer model for estimating vegetation LAI and chlorophyll in a heterogeneous grassland

Roshanak Darvishzadeh a,⁎, Andrew Skidmore a, Martin Schlerf a, Clement Atzberger b

a International Institute for Geo-information Science and Earth Observation (ITC), Hengelosestraat 99, P.O. Box 6, 7500 AA Enschede, The Netherlands
b Joint Research Centre of the European Commission, TP 266, Via Enrico Fermi 1, 21020 Ispra (VA), Italy

Received 12 October 2007; received in revised form 29 November 2007; accepted 15 December 2007

Abstract

Radiative transfer models have seldom been applied for studying heterogeneous grassland canopies. Here, the potential of radiative transfer modeling to predict LAI and leaf and canopy chlorophyll contents in a heterogeneous Mediterranean grassland is investigated. The widely used PROSAIL model was inverted with canopy spectral reflectance measurements by means of a look-up table (LUT). Canopy spectral measurements were acquired in the field using a GER 3700 spectroradiometer, along with simultaneous in situ measurements of LAI and leaf chlorophyll content. We tested the impact of using multiple solutions, stratification (according to species richness), and spectral subsetting on parameter retrieval. To assess the performance of the model inversion, the normalized RMSE and \( R^2 \) between independent in situ measurements and estimated parameters were used. Of the three investigated plant characteristics, canopy chlorophyll content was estimated with the highest accuracy \( R^2 = 0.70, \text{ NRMSE} = 0.18 \). Leaf chlorophyll content, on the other hand, could not be estimated with acceptable accuracy, while LAI was estimated with intermediate accuracy \( R^2 = 0.59, \text{ NRMSE} = 0.18 \). When only sample plots with up to two species were considered \( (n = 107) \), the estimation accuracy for all investigated variables (LAI, canopy chlorophyll content and leaf chlorophyll content) increased \( (\text{NRMSE} = 0.14, 0.16, 0.19, \text{ respectively}) \). This shows the limits of the PROSAIL radiative transfer model in the case of very heterogeneous conditions. We also found that a carefully selected spectral subset contains sufficient information for a successful model inversion. Our results confirm the potential of model inversion for estimating vegetation biophysical parameters at the canopy scale in (moderately) heterogeneous grasslands using hyperspectral measurements.

© 2008 Elsevier Inc. All rights reserved.

Keywords: Hyperspectral; Grassland; LAI; Leaf chlorophyll; Canopy chlorophyll; Model inversion; Radiative transfer model; LUT

1. Introduction

Accurate quantitative estimation of vegetation biochemical and biophysical variables is useful for a large variety of agricultural, ecological, and meteorological applications (Asner, 1998; Houborg et al., 2007). The spatial and temporal distribution of vegetation biochemical and biophysical variables are important inputs into models quantifying the exchange of energy and matter between the land surface and the atmosphere. Among the many vegetation characteristics, leaf area index (LAI), leaf chlorophyll content (LCC) and canopy chlorophyll content (CCC) are of prime importance (Bacour et al., 2006; Houborg et al., 2007). LAI, defined here as a one-sided leaf area divided by unit of horizontal surface area, is a key structural characteristic of vegetation because of the role of green leaves in controlling many biological and physical processes in plant canopies. Leaf chlorophyll content and canopy chlorophyll content (the latter defined here as the product of LAI and leaf chlorophyll content) contribute to verifying vegetation physiological status and health, and have been found useful for detecting vegetation stress, photosynthetic capacity, and productivity (Boegh et al., 2002; Carter, 1994).

There are two common approaches to estimating vegetation parameters (including LAI and chlorophyll) from remotely sensed data. In the empirical/statistical approach, statistical techniques are used to obtain a correlation between the target variable (e.g., LAI measured in situ) and its spectral reflectance or some vegetation indices. The derived statistical relationships are recognized as being sensor-specific and dependent on site
and sampling condition, and are expected to change in space and time (Colombo et al., 2003; Meroni et al., 2004). The physical approach, on the other hand, involves using radiative transfer models. This approach assumes that the radiative transfer model accurately describes the spectral variation of canopy reflectance, as a function of canopy, leaf and soil background characteristics, using physical laws (Goel, 1989; Meroni et al., 2004). As radiative transfer models are able to explain the transfer and interaction of radiation inside the canopy based on physical laws, they offer an explicit connection between the vegetation biophysical and biochemical variables and the canopy reflectance (Houborg et al., 2007).

To actually use physically based models for retrieving vegetation characteristics from observed reflectance data, they must be inverted (Kimes et al., 1998). Different inversion techniques have been proposed for physical models, including numerical optimization methods (Jacquemoud et al., 2000; Jacquemoud et al., 1995; Meroni et al., 2004), look-up table (LUT) approaches (Combal et al., 2002; Combal et al., 2003; Gastellu-Etchegorry et al., 2003; Weiss et al., 2000), artificial neural networks (Schlerf and Atzberger, 2006; Walthall et al., 2004; Weiss and Baret, 1999) and, very recently, support vector machines regression (Durba et al., 2007). In the iterative optimization approach, a stable and optimum inversion is not guaranteed, as the search algorithm may get trapped in local minima before reaching the global minimum. Moreover, the technique is computationally intensive, in particular when using complex radiative transfer models. This makes the retrieval of biophysical variables impossible for large geographic areas (Houborg et al., 2007). LUT and neural network approaches reduce the huge computational demand of the traditional optimization approach (Kimes et al., 2000; Liang, 2004). However, for proper training (artificial neural networks) and representation (LUT), they rely on a large database of simulated canopy reflectance spectra to achieve a high degree of accuracy. This increases the computational time for identifying the most appropriate LUT entry (Liang, 2004) and the time required for training the artificial neural network. The interested reader may refer to Kimes et al. (2000) and Liang (2004) for more detailed discussions regarding the advantages and disadvantages of the three inversion methods.

A drawback in using physically based models is the ill-posed nature of model inversion (Atzberger, 2004; Combal et al., 2002), meaning that the inverse solution is not always unique as various combinations of canopy parameters may yield almost similar spectra (Weiss and Baret, 1999). To overcome this problem, some restriction of the inverse process may be required to constrain the inversion process. This involves the use of prior knowledge about model parameters (Combal et al., 2002; Lavergne et al., 2007), the use of information provided by the temporal course of key canopy parameters (CROMA, 2000), and/or the analysis of color textures and object signatures (Atzberger, 2004).

Significant efforts to estimate and quantify vegetation properties using radiative transfer models have been carried out in the last two decades. Several studies have been successfully conducted covering different vegetation types and remote sensing data: on global data sets (Bacour et al., 2006; Baret et al., 2007), on agricultural crops (Atzberger, 2004; Jacquemoud et al., 2000; Jacquemoud et al., 1995), on semi-arid regions (Qi et al., 2000), and on forests (Gemmell et al., 2002; Kötz et al., 2004; Meroni et al., 2004; Schlerf and Atzberger, 2006; Zarco-Tejada et al., 2004a; Zarco-Tejada et al., 2004b). Many other studies have analyzed simulated data (Gong et al., 1999; Weiss et al., 2000). Despite these efforts, the review of the literature reveals that there is a gap in estimating vegetation biophysical and biochemical variables for heterogeneous areas such as heterogeneous grasslands with combinations of different grass species. Grasslands are of vital ecological importance to the conservation of biodiversity in managed agricultural landscapes. Furthermore, studies that use hyperspectral measurements and that include validation with large numbers of ground truth data for heterogeneous grasslands are extremely rare.

The main objective of this study was to estimate and predict canopy characteristics such as LAI and chlorophyll content in a heterogeneous Mediterranean grassland by inverting the canopy radiative transfer model PROSAIL (Jacquemoud and Baret, 1990; Verhoef, 1984; Verhoef, 1985). The study is based on canopy spectral reflectance measured during a field campaign in the summer of 2005 in Majella National Park in Italy. A LUT-based inversion algorithm has been used, accounting for available prior information relating to the distribution (probable range) of several vegetation characteristics. The suitability of the methods is analyzed in terms of prediction accuracy for estimating LAI, leaf chlorophyll content and canopy chlorophyll content.

2. Materials and 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 (Fig. 1). 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). Geologically, the region is made up of calcareous rocks, which date back to the Jurassic period. The flora of the park includes more than 1800 plant species, which constitute approximately one third of the entire flora of Italy (Cimini, 2005).

Abandoned agricultural areas and settlements in Majella are returning to oak (Quercus pubescens) woodlands at the lower altitude (400 m to 600 m) and beech (Fagus sylvatica) forests at higher altitudes (1200 m to 1800 m). Between these two formations is a landscape composed of shrubby bushes, patches of grass/herb vegetation, and bare rock outcrops. The dominant grass and herb species include Brachypodium pinnatum, Briza media, Bromus erectus, Festuca sp., Helichrysum italicum, Galium verum, Trifolium pratense, Plantago lanceolata, Sanguisorba officinalis and Ononis spinosa (Cho, 2007).

Stratified random sampling was adopted in this study. For this purpose, the area was stratified into grassland, forest, shrubland and bare rock outcrops, using a land cover map provided by the
management of Majella National Park. We distinguished four main phytosociological classes of varying areas within the grasslands: semi-natural/farmlands, grazed/periodically flooded areas, open garrigues and abandoned farmlands. Coordinates (x y) were randomly generated in the grassland stratum to select plots. A total of 45 quadratic plots of 30 m side length were generated and a GPS was used to locate their position in the field. To increase the number of samples in the time available, within each plot four to five randomly selected subplots were identified. This resulted in a total of 185 subplots being sampled (from the original 191 subplots, six subplots showed poor quality and had to be discarded). Each subplot covered 1 m × 1 m (average vegetation height, 28 cm), with different species compositions and relative abundances while the within-subplot variability was small. The species varied in terms of leaf shape, leaf size, the amount of leaves and their typical angle distribution. The within-subplot variability of SPAD measurements also indicated some variation in chlorophyll contents, albeit this has not been quantified within the present study.

2.2. Canopy spectral measurements

Fifteen replicates of canopy spectral measurements were taken from each subplot, using a GER 3700 (Geophysical and Environmental Research Corporation, Buffalo, New York) spectroradiometer. The GER 3700 is a three dispersion grating spectroradiometer using Si and PbS detectors with a single field of view. The wavelength range is 350 nm to 2500 nm, with a spectral sampling of 1.5 nm in the 350 nm to 1050 nm range, 6.2 nm in the 1050 nm to 1900 nm range, and 9.5 nm in the 1900 nm to 2500 nm range. The spectral resolution (band pass) is 3 nm, 11 nm and 16 nm in the 350 nm to 1050 nm range, 1050 nm to 1900 nm range, and 1900 nm to 2500 nm range, respectively.

The fiber optic, with a field of view of 25°, was hand-held approximately 1 m above the ground at nadir position. The ground area observed by the sensor 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 the measurement noise to be averaged out. 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. The measurement setup ensured that the ratio of direct to diffuse incoming solar radiation was approximately constant. Hence, no correction for this possibly perturbing factor has been applied.

The spectral reflectance of bare soils was acquired from a few subplots with no vegetation and their average was calculated. Mean reflectance spectra are shown in Fig. 2. We assumed that the measured soil optical properties were representative for the study area.

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). Measurements were taken either under clear skies with low solar elevation (i.e., within the 2 h following sunrise or preceding sunset) or under overcast conditions. Care was taken to measure LAI on the same day as the canopy spectral measurements were made. To prevent direct sunlight on the sensor, samples of below- and above-canopy radiation were made with the sun behind the operator and using a view restrictor of 45°. For each subplot, reference sample of above-canopy radiation was taken by measuring incoming radiation above the grass subplot. Next, five below-canopy samples were collected and used to calculate the average LAI.

LAI measured using the LAI-2000 corresponds to plant area index (PAI), including photosynthetic and non-photosynthetic components (Chen et al., 1997). In our study, non-photosynthetic components were almost non-existent. Despite the non-random distribution of grass leaves, no corrections for clumping were applied. Therefore, the LAI used here corresponds to effective PAI, and in the following sections these measurements are abbreviated as LAI.

The statistics of the 185 samples comprising different grass species are summarized in Table 1. The table reveals a large range of LAI values, which enables the approach to be validated under contrasting conditions.

2.4. Chlorophyll measurements

In each 1 m×1 m subplot, a SPAD-502 Leaf Chlorophyll Meter (MINOLTA, Inc.) was used to assess leaf chlorophyll content. SPAD values express relative amounts of chlorophyll in leaves by measuring transmittance in the red (650 nm) and near-infrared (920 nm) wavelength regions (Minolta, 2003). SPAD measurements give a unitless but highly reproducible measure, which is well correlated with leaf chlorophyll concentration, and is commonly used to characterize chlorophyll concentration in many plant species (Campbell et al., 1990; Haboudane et al., 2002; Jongsaap and Booij, 2004; Nakano et al., 2006). 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. These averaged SPAD readings (unitless) were converted into leaf chlorophyll content (µg cm⁻²) by means of an empirical calibration function provided by Markwell et al. (1995). Although the Markwell function refers to soybean and corn leaves, the same authors have demonstrated that they can also be applied to other plant species. Hence, we renounced to establish specific calibration functions for the grass species since each sample plot consist of several species. The total canopy chlorophyll content (g m⁻²) for each subplot was obtained by multiplying the leaf chlorophyll content by the corresponding LAI.

2.5. Pre-processing of spectra

To minimize noise in the measured reflectance spectra, the 15 spectra of each sample plot were averaged. Bands with a wavelength less than 400 nm and more than 2400 nm displayed very high levels of noise and were excluded. The resulting 584 wavebands were used for the analysis. A moving Savitzky–Golay filter (Savitzky and Golay, 1964) with a frame size of 15 data points (2nd degree polynomial) was applied to the averaged reflectance measurements to further smooth the spectra. The analysis and processing was carried out using MATLAB 7.1 (Mathworks, 2007). In total, 185 canopy reflectance spectra were obtained. The average reflectance spectra of all grass subplots and the spectral variability of the measurements are shown in Fig. 2.

2.6. The PROSAIL radiative transfer model

The widely used PROSAIL radiative transfer model, which is a combination of the SAILH canopy reflectance model (Verhoef, 1984; Verhoef, 1985; Kuusk, 1991) and the PROSPECT leaf optical properties model (Jacquemoud and Baret, 1990), was used to retrieve the LAI and leaf and canopy chlorophyll contents. Both submodels are relatively simple and need only a limited number of input parameters, with reasonable computation time. By inverting the coupled models, both the leaf and canopy parameters can be estimated.

The PROSPECT model (Fourty et al., 1996; Jacquemoud and Baret, 1990; Jacquemoud et al., 1996) calculates the leaf hemispherical transmittance and reflectance as a function of four input parameters, i.e., the leaf structural parameter, N (unitless); the leaf chlorophyll a+b concentration, LCC (µg cm⁻²); the dry matter content, Cm (g cm⁻²); and the equivalent water thickness, Cw (g cm⁻²) (Jacquemoud et al., 2000). The spectral leaf optical properties (reflectance and transmittance) calculated by PROSPECT are inputs into the SAILH canopy reflectance model. This model (Verhoef, 1984; Verhoef, 1985) is a one-dimensional bidirectional turbid medium radiative transfer model that has been later modified to take into account the hot spot effect in plant canopy reflectance (Kuusk, 1991). Turbid medium defines the canopy as a horizontally homogenous and semi-infinite layer that consists of small vegetation elements that act as absorbing and scattering particles of a given geometry and density. Consequently, the model is best adopted for use in homogeneous vegetation canopies (Meroni et al., 2004; Schlfer and Atzberger, 2006; Verhoef, 1984). Apart from the leaf reflectance and transmittance, the SAILH model requires eight input parameters to produce the top-of-canopy bidirectional reflectance. These are sun zenith angle, \( t_s \) (deg); sensor viewing angle, \( t_v \) (deg); azimuth angle, \( \phi \) (deg); fraction of diffuse incoming solar

<table>
<thead>
<tr>
<th>Measured variables</th>
<th>Min</th>
<th>Mean</th>
<th>Max</th>
<th>StDev</th>
<th>Range</th>
<th>Variation coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>LAI (m² m⁻²)</td>
<td>0.39</td>
<td>2.81</td>
<td>7.34</td>
<td>1.50</td>
<td>6.95</td>
<td>0.53</td>
</tr>
<tr>
<td>SPAD (unitless)</td>
<td>22.4</td>
<td>32.70</td>
<td>45</td>
<td>4.35</td>
<td>22.6</td>
<td>0.13</td>
</tr>
<tr>
<td>LCC (µg cm⁻²)</td>
<td>17.1</td>
<td>30.07</td>
<td>49.66</td>
<td>6.12</td>
<td>32.55</td>
<td>0.20</td>
</tr>
<tr>
<td>CCC (g m⁻²)</td>
<td>0.1</td>
<td>0.87</td>
<td>2.7</td>
<td>0.55</td>
<td>2.56</td>
<td>0.63</td>
</tr>
<tr>
<td>Dominant species number</td>
<td>1</td>
<td>2.34</td>
<td>4</td>
<td>0.87</td>
<td>3.07</td>
<td>0.37</td>
</tr>
</tbody>
</table>

SPAD is the average SPAD reading for 30 randomly selected leaves in each subplot; LCC is the leaf chlorophyll content; CCC is the canopy chlorophyll content.

Table 1

Summary statistics of the measured biophysical and biochemical variables of grassland sample plots (n=185)
Table 2
Specific ranges for eight input parameters used for generating the LUT, using forward calculation, of the PROSAIL model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Abbr. in model</th>
<th>Unit</th>
<th>Minimum value</th>
<th>Maximum value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leaf area index</td>
<td>LAI</td>
<td>m² m⁻²</td>
<td>0.3</td>
<td>7.5</td>
</tr>
<tr>
<td>Mean leaf inclination angle</td>
<td>ALA</td>
<td>Deg</td>
<td>40</td>
<td>70</td>
</tr>
<tr>
<td>Leaf chlorophyll content</td>
<td>LCC</td>
<td>μg cm⁻²</td>
<td>15</td>
<td>55</td>
</tr>
<tr>
<td>Leaf structural parameter</td>
<td>N</td>
<td>No dimension</td>
<td>1.5</td>
<td>1.9</td>
</tr>
<tr>
<td>Dry matter content</td>
<td>C_m</td>
<td>g cm⁻²</td>
<td>0.005</td>
<td>0.01</td>
</tr>
<tr>
<td>Equivalent water thickness</td>
<td>C_w</td>
<td>g cm⁻²</td>
<td>0.01</td>
<td>0.02</td>
</tr>
<tr>
<td>Hot spot size parameter</td>
<td>hot</td>
<td>m m⁻¹</td>
<td>0.05</td>
<td>0.1</td>
</tr>
<tr>
<td>Soil brightness parameter</td>
<td>scale</td>
<td>No dimension</td>
<td>0.5</td>
<td>1.5</td>
</tr>
</tbody>
</table>

Within the specified ranges, parameter values were drawn randomly (uniform distributions).

* The minimum and maximum values are selected based on the prior knowledge from the field.

2.7. The look-up table (LUT) inversion

Perhaps the simplest method of solving the inversion of a radiative transfer model is by using a LUT. LUTs offer an interesting alternative to numerical optimization and neural network methods because they permit a global search (avoiding local minima) while showing less unexpected behavior when the spectral characteristics of the targets are not well represented by the modeled spectra (Schlerf and Atzberger, 2006). A LUT is built in advance of the actual inversion through forward calculations using a radiative transfer model. For the inversion, only search operations are needed to identify the parameter combinations that yield the best fit between measured and LUT spectra. However, to achieve high accuracy for the estimated parameters, the dimension of the LUT must be sufficiently large (Combal et al., 2002; Tang et al., 2006; Weiss et al., 2000).

To build the LUT, 100,000 parameter combinations were randomly generated (uniform distributions) and used in the forward calculation of the PROSAIL model. We also tested normally distributed random parameters and found no significant differences (not shown). The ranges (minimum and maximum) for each of the eight “free” model parameters are reported in Table 2. To prevent too-wide parameter spaces and to reduce the size of the parameter spacing, the maximum and minimum values of LAI, LCC, and ALA (recorded along with LAI using LAI-2000 instrument) were fixed based on the prior knowledge from the field data collection (Combal et al., 2003). Parameters difficult to measure (e.g., N, C_m, C_w) are often fixed to nominal values (e.g. Chaurasia and Dadhwal, 2004; Haboudane et al., 2004; Houborg et al., 2007; le Maire et al., 2004). For the leaf structural parameter N of PROSPECT, Haboudane et al. (2004) and Houborg et al. (2007) have used a fixed value of 1.55 for various crops, including corn, soybean, wheat and spring barley. Jacquemoud et al. (2000) have used a fixed value of N=1.7 for soybean. Atzberger et al. (2003) have used a range of N=2±0.34 for wheat crop. Since grasses have relatively thin leaves, we used for the N parameter a range from 1.5 to 1.9. The ranges of other input parameters (C_w, C_m, hot and scale) were selected similarly in agreement with the existing literature (Atzberger, 2004; Cho, 2007; Combal et al., 2003; Haboudane et al., 2004; le Maire et al., 2004; Schlerf and Atzberger, 2006). We used the average bare soil reflectance spectrum that was measured in the study area to represent soil optical properties (Fig. 2). Since the spectral measurements were done around noon with the sensor looking at nadir position, the sensor viewing angle (t_v), the relative azimuth angle (phi) and the average sun zenith angle (t_s) were fixed at 0°, 0° and 30°, respectively, representing the geometry of the measurement setup. With respect to the fraction of diffuse incoming solar radiation, sky radiance (skyl), a fixed value of 0.1 across all wavelengths has been used, as in many similar studies (Cho, 2007; Schlerf and Atzberger, 2006). Hence, we neglect that the amount of diffuse sky light depends on atmospheric conditions, solar zenith angle and furthermore is wavelength dependent. This simplification seems justified, however, by the fact that sky radiance has only a very small influence on canopy reflectance (Clevers and Verhoef, 1991) and by the lack of on-site measurements of skyl.

To find the solution to the inverse problem for a given canopy spectrum for each modeled reflectance spectrum of the LUT, the RMSE between measured and modeled spectra (RMSE_m) is calculated according to:

\[
\text{RMSE}_m = \sqrt{\frac{\sum_{i=1}^{n} (R_{\text{measured}_i} - R_{\text{LUT}_i})^2}{n}}
\]

where \(R_{\text{measured}_i}\) is a measured reflectance at wavelength \(\lambda_i\) and \(R_{\text{LUT}_i}\) is a modeled reflectance at wavelength \(\lambda_i\) in the LUT, and \(n\) is the number of wavelengths. Traditionally, the solution is regarded as the set of input parameters corresponding to the reflectance in the LUT that provides the smallest RMSE_m. However, this solution is not always the optimal solution since it may not be unique (ill-posed problem). To overcome this problem and to enhance the consistency of the estimated variables, we also investigated the use of some other statistical indicators, such as the mean and median from the best 10, 20, 40 and 100 simulations.

An appropriate band selection – or alternatively, the weighting of different spectral bands – is known to improve radiative transfer model inversion and prevents bias in the estimation of the variables of interest (Bacour et al., 2001; Meroni et al., 2004;
Schlerf and Atzberger, 2006; Lavergne et al., 2007). This is particularly the case if hyperspectral data with wavelengths that are either noisy or not well modeled by the radiative transfer model being inverted. Nevertheless, the selection of an optimal spectral subset/weighting of spectral bands is not a trivial problem and is still an open issue within the remote sensing community (Meroni et al., 2004; Lavergne et al., 2007).

To investigate the role of heterogeneity (number of dominant plant species within a subplot) in the estimation of grass variables, we also stratified the data based on dominant species composition.

3. Results

3.1. Grass characteristics

Each subplot varied in species composition and biophysical/biochemical characteristics (Table 1). Consequently, spectral reflectance measurements showed considerable variability (Fig. 2). Linear correlation between the canopy characteristics confirmed independence between leaf chlorophyll content and LAI, while LAI and canopy chlorophyll content were highly correlated (Table 3). This is explained by larger coefficient of variation of LAI compared to leaf chlorophyll content (Table 1).

3.2. Inversion results based on the smallest RMSE criterion

To find the solution to the inverse problem, the LUT is sorted according to the cost function (RMSE) and the set of variables providing the minimum RMSE is considered as the solution. Fig. 3 illustrates measured and simulated canopy reflectance spectra found in this way for three subplots with contrasting LAI values.

In the examples, the simulated reflectances were in relatively good agreement with the measured reflectances for canopies with different LAI values. From the analysis of 185 canopy reflectance spectra, we found that medium range LAI sample plots were best modeled by PROSAIL (lowest RMSE, between measured and modeled spectra). In general, differences between measured and modeled spectral reflectances were inconsistent, even among canopies with a single species. Fig. 4 demonstrates the average absolute error (AAE) between measured and best-fit spectra as a function of wavelengths. The figure shows that the AAE in some regions is relatively high, especially for the water vapor absorption regions (1135 nm to 1400 nm, and 1820 nm to 1940 nm). The canopy reflectance in these regions is either not well measured or not well modeled by PROSAIL (see Section 3.5 for spectral subsetting).

The relation between the measured and estimated grass variables based on the smallest RMSE criterion is demonstrated in Fig. 5. The $R^2$ and the normalized RMSE (NRMSE = RMSE/range) (Atzberger, 1997; Combal et al., 2003) between measured and estimated leaf chlorophyll content indicate poor relationships. LAI and canopy chlorophyll content were estimated with much higher accuracy. As the canopy reflectance is modulated mostly by LAI and the integrated chlorophyll content of the canopy (hence canopy chlorophyll content), which both showed

<table>
<thead>
<tr>
<th>Grass canopy characteristic</th>
<th>LAI (m² m⁻²)</th>
<th>LCC (µg cm⁻²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LCC (µg cm⁻²)</td>
<td>0.24</td>
<td>–</td>
</tr>
<tr>
<td>CCC (g m⁻²)</td>
<td>0.94</td>
<td>0.50</td>
</tr>
</tbody>
</table>

LCC is the leaf chlorophyll content; and CCC is the canopy chlorophyll content. Correlation coefficients significant at $p \leq 0.001$. 

Fig. 3. Measured and simulated grass canopy reflectance spectra of three sample plots with LAI equal to 0.9 (left), 2.82 (center), and 6.16 (right), respectively.

Fig. 4. The average absolute error (AAE) between measured and best-fit reflectance spectra as a function of wavelengths. The AAE has been calculated from the 185 measured canopy spectra against the best fitting look-up table (LUT) spectra.

Fig. 5. The $R^2$ and the normalized RMSE (NRMSE = RMSE/range) between measured and estimated leaf chlorophyll content.
considerable variability (Table 1), the poor retrieval of the leaf chlorophyll content was expected (lower variability).

Studying the histograms of the other 6 retrieved parameters ($C_m$, $C_w$, scale, ALA, hot and $N$) revealed that several (160 out of 185) samples’ plots reached the upper/lower boundary of at least one model parameter. As the parameter ranges were relatively large and consistent with available field observations (Table 2), we believe that some wavelengths are either badly measured or not well modeled by the combined SAILH and PROSPECT canopy reflectance model (Schlerf & Atzberger 2006).

3.3. Inversion results based on multiple solutions

For each measured canopy spectra, the LUT was sorted from minimum to maximum RMSE$_r$ (Eq. (1)). Instead of taking simply the PROSAIL parameter corresponding to the lowest RMSE$_r$ (Section 3.2), we alternatively tested to consider the best 10, 20, 40 and 100 LUT entries as the final solution. The importance of considering multiple solutions rather than the single LUT solution with minimum RMSE is seen from Table 4. It demonstrates how different solutions affect the accuracy of the estimated variables. We used one-way ANOVA (analysis of variance) to evaluate the existence of significant differences in the mean $R^2$ between the median/mean for the three biophysical grass variables. The test was conducted for the four multiple solutions (i.e., the first 10, the first 20, the first 40 and the first 100 best fits). The results show that, generally, there are no significant differences between the statistical parameters used for any number of solutions ($p > 0.05$). Nevertheless, throughout the rest of this study we considered the first 100 solutions as the best measures for estimating the grass variables.

3.4. Inversion results based on stratification of heterogeneity

According to the number of dominant species, the subplots were divided into seven data sets (Table 5). The statistical

Table 4  
$R^2$, RMSE and normalized RMSE between measured and estimated grass characteristics ($n = 185$)

<table>
<thead>
<tr>
<th>Nr. of solutions</th>
<th>Statistical parameter</th>
<th>$R^2$</th>
<th>RMSE</th>
<th>NRMSE</th>
<th>$R^2$</th>
<th>RMSE</th>
<th>NRMSE</th>
<th>$R^2$</th>
<th>RMSE</th>
<th>NRMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Best fitting spectra</td>
<td>/</td>
<td>0.59</td>
<td>1.28</td>
<td>0.18</td>
<td>0.27</td>
<td>6.8</td>
<td>0.21</td>
<td>0.70</td>
<td>0.45</td>
<td>0.18</td>
</tr>
<tr>
<td>First 10</td>
<td>Median</td>
<td>0.61</td>
<td>1.21</td>
<td>0.17</td>
<td>0.31</td>
<td>5.7</td>
<td>0.18</td>
<td>0.71</td>
<td>0.43</td>
<td>0.17</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td>0.62</td>
<td>1.18</td>
<td>0.17</td>
<td>0.30</td>
<td>6.0</td>
<td>0.18</td>
<td>0.71</td>
<td>0.43</td>
<td>0.17</td>
</tr>
<tr>
<td>First 20</td>
<td>Median</td>
<td>0.61</td>
<td>1.18</td>
<td>0.17</td>
<td>0.34</td>
<td>5.7</td>
<td>0.18</td>
<td>0.71</td>
<td>0.44</td>
<td>0.17</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td>0.62</td>
<td>1.15</td>
<td>0.17</td>
<td>0.33</td>
<td>5.7</td>
<td>0.18</td>
<td>0.72</td>
<td>0.43</td>
<td>0.17</td>
</tr>
<tr>
<td>First 40</td>
<td>Median</td>
<td>0.60</td>
<td>1.21</td>
<td>0.17</td>
<td>0.24</td>
<td>7.5</td>
<td>0.23</td>
<td>0.70</td>
<td>0.45</td>
<td>0.18</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td>0.61</td>
<td>1.18</td>
<td>0.17</td>
<td>0.27</td>
<td>6.6</td>
<td>0.20</td>
<td>0.70</td>
<td>0.44</td>
<td>0.17</td>
</tr>
<tr>
<td>First 100</td>
<td>Median</td>
<td>0.64</td>
<td>1.1</td>
<td>0.16</td>
<td>0.35</td>
<td>5.3</td>
<td>0.16</td>
<td>0.72</td>
<td>0.42</td>
<td>0.16</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td>0.63</td>
<td>1.1</td>
<td>0.16</td>
<td>0.35</td>
<td>5.4</td>
<td>0.17</td>
<td>0.72</td>
<td>0.42</td>
<td>0.16</td>
</tr>
</tbody>
</table>

The standard LUT solution is indicated as “best fitting spectra”. The grass characteristics were also retrieved considering the first 10, 20, 40 and 100 solutions. In these cases, the median and mean were investigated. LCC is the leaf chlorophyll content and CCC is the canopy chlorophyll content.
analysis was done separately for each of the seven data sets. We considered the “best fitting spectra” and the first 100 solutions for estimating the grass variables (see Section 3.3). Table 5 shows the results of this stratification.

It can be seen from Table 5 that stratification based on the number of species has a strong influence on the estimation accuracy for the grass variables, in particular for LAI and canopy chlorophyll content. For leaf chlorophyll content, on the other hand, no trend could be observed. The estimation accuracy for LAI increased considerably, from $R^2 = 0.59$, NRMSE = 0.18 for up to four species (i.e., all subplots) to $R^2 = 0.81$, NRMSE = 0.11 for one species. As regards canopy chlorophyll content, the effect of species reduction was weaker (up to four species: $R^2 = 0.70$, NRMSE = 0.18; one species $R^2 = 0.78$, NRMSE = 0.16). The table also indicates that the inversion of the PROSAIL model enables grass variables to be estimated with relatively good accuracy if only canopies with up to two species are considered (LAI: $R^2 = 0.71$, NRMSE = 0.14; CCC: $R^2 = 0.80$, NRMSE = 0.16). Measured and estimated grass variables of subplots with up to two species are shown in Fig. 6. The result suggests that the PROSAIL model is not well adapted to multi-species grasslands.

Table 5

<table>
<thead>
<tr>
<th>Stratification/dominant species</th>
<th>Statistical parameter</th>
<th>LAI (m² m⁻²)</th>
<th>LCC (µg cm⁻²)</th>
<th>CCC (g m⁻²)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$R^2$</td>
<td>RMSE</td>
<td>NRMSE</td>
<td>$R^2$</td>
</tr>
<tr>
<td>One species ($n=32$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Best fitting spectra</td>
<td>0.81</td>
<td>0.76</td>
<td>0.11</td>
<td>0.21</td>
</tr>
<tr>
<td>Median of 100</td>
<td>0.85</td>
<td>0.68</td>
<td>0.10</td>
<td>0.19</td>
</tr>
<tr>
<td>Mean of 100</td>
<td>0.85</td>
<td>0.67</td>
<td>0.10</td>
<td>0.17</td>
</tr>
<tr>
<td>Two species ($n=75$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Best fitting spectra</td>
<td>0.69</td>
<td>1.1</td>
<td>0.17</td>
<td>0.37</td>
</tr>
<tr>
<td>Median of 100</td>
<td>0.69</td>
<td>1.1</td>
<td>0.17</td>
<td>0.41</td>
</tr>
<tr>
<td>Mean of 100</td>
<td>0.69</td>
<td>1.1</td>
<td>0.17</td>
<td>0.40</td>
</tr>
<tr>
<td>Three species ($n=59$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Best fitting spectra</td>
<td>0.55</td>
<td>1.6</td>
<td>0.33</td>
<td>0.13</td>
</tr>
<tr>
<td>Median of 100</td>
<td>0.56</td>
<td>1.2</td>
<td>0.25</td>
<td>0.32</td>
</tr>
<tr>
<td>Mean of 100</td>
<td>0.56</td>
<td>1.2</td>
<td>0.25</td>
<td>0.33</td>
</tr>
<tr>
<td>Four species ($n=19$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Best fitting spectra</td>
<td>0.37</td>
<td>1.49</td>
<td>0.37</td>
<td>0.26</td>
</tr>
<tr>
<td>Median of 100</td>
<td>0.38</td>
<td>1.3</td>
<td>0.32</td>
<td>0.45</td>
</tr>
<tr>
<td>Mean of 100</td>
<td>0.38</td>
<td>1.3</td>
<td>0.32</td>
<td>0.45</td>
</tr>
<tr>
<td>Up to two species ($n=107$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Best fitting spectra</td>
<td>0.71</td>
<td>0.99</td>
<td>0.14</td>
<td>0.40</td>
</tr>
<tr>
<td>Median of 100</td>
<td>0.72</td>
<td>0.97</td>
<td>0.14</td>
<td>0.41</td>
</tr>
<tr>
<td>Mean of 100</td>
<td>0.71</td>
<td>0.97</td>
<td>0.14</td>
<td>0.40</td>
</tr>
<tr>
<td>Up to three species ($n=166$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Best fitting spectra</td>
<td>0.62</td>
<td>1.26</td>
<td>0.18</td>
<td>0.26</td>
</tr>
<tr>
<td>Median of 100</td>
<td>0.66</td>
<td>1.1</td>
<td>0.16</td>
<td>0.34</td>
</tr>
<tr>
<td>Mean of 100</td>
<td>0.66</td>
<td>1.1</td>
<td>0.16</td>
<td>0.33</td>
</tr>
<tr>
<td>Up to four species ($n=185$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(all subplots)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Best fitting spectra</td>
<td>0.59</td>
<td>1.28</td>
<td>0.18</td>
<td>0.27</td>
</tr>
<tr>
<td>Median of 100</td>
<td>0.64</td>
<td>1.1</td>
<td>0.16</td>
<td>0.35</td>
</tr>
<tr>
<td>Mean of 100</td>
<td>0.63</td>
<td>1.1</td>
<td>0.16</td>
<td>0.35</td>
</tr>
</tbody>
</table>

LCC is the leaf chlorophyll content and CCC is the canopy chlorophyll content.

Fig. 6. Measured versus estimated grass variables based on subplots with up to two species in Majella National Park, Italy ($n=107$). Left: LAI, center: leaf chlorophyll content, and right: canopy chlorophyll content.
Fig. 6 shows that LAI values up to 4 were better estimated than higher LAI values. It seems that canopies with LAI values greater than 4 were somehow overestimated, whereas canopies with LAI values less than 4 were slightly underestimated. This is seen in the linear relation between measured and retrieved LAI that gave a slope of 0.96 and an intercept of 0.28, and can be considered both as model error and bias in the reference measurements. Comparison of Figs. 5 and 6 reveals that larger scatter was introduced in the estimated grass variables by increasing the number of species. This again confirms that the PROSAIL model is best adapted to canopies with few species (for which it was conceived). No relationships were observed between RMSEr (best fit) and the residuals of the variables ($R^2 = 0.008, 0.057$ and $0.007$ for LAI, leaf chlorophyll content and canopy chlorophyll content, respectively), indicating that the model was properly fitted.

3.5. Inversion results based on spectral sampling

In this study, three spectral subsets have been used. Two subsets were prepared based on the results of a previous study by Darvishzadeh et al. (in press) that included the selection of wavelengths through stepwise multiple linear regression (Table 6), as well as the use of a subset of wavelengths closely related to vegetation parameters identified from literature (Cho et al., 2007; Darvishzadeh et al., in press) (Table 7). The third subset was constructed based on the average absolute error (AAE) between the measured and best-fit reflectance spectra (Fig. 4). We considered the bands with an AAE greater or equal to 0.02 as wavelengths with high errors. These bands were systematically excluded (one by one) in the inversion process, and each time the AAE between the measured and best-fit reflectance spectra was calculated until the remaining wavelengths were left with an AAE smaller than 0.02. Fig. 7 shows the distribution of the spectral regions with AAE greater or equal to 0.02 that were removed from the existing wavelengths. The remaining wavebands (384) were considered as our third spectral subset. Spectral subsets were prepared for the entire field data set ($n=185$).

The role of the spectral subsets in the estimation of grass variables was again evaluated on the basis of the $R^2$ and the (normalized) RMSE between the measured and estimated grass variables. The results showed that, compared with using all wavebands, employing spectral subset I (Table 6) gave significantly larger errors for LAI and leaf chlorophyll content (Table 8). In contrast, by employing subset II (Table 7) and subset III (Fig. 7) instead of the full spectral resolution (Table 8), the relationships between measured and estimated LAI (and leaf chlorophyll content) were almost similar.

However, it seemed that the estimation accuracies between measured and estimated canopy chlorophyll content improved using all three subsets (Table 8). This shows that the full hyperspectral resolution is not automatically more advantageous than a carefully designed multi-spectral sensor (e.g., Fourt and Baret, 1997). For example, some bands may contain (excessively) high noise levels and therefore damage the results. The same holds for bands that, for various reasons, are not well modeled by the radiative transfer model (see Fig. 4, the water absorption regions).

4. Discussion

The canopy-integrated chlorophyll content ($\text{LAI} \times \text{leaf chlorophyll content}$) strongly reflects the variability of LAI as the leaf chlorophyll content was relatively stable (Table 1). The
LCC is the leaf chlorophyll content, and CCC is the canopy chlorophyll content.

In general, the relationships between measured and estimated leaf chlorophyll content were poor in all inversion processes. This confirms other studies revealing similar difficulties in estimating leaf chlorophyll (Baret and Jacquemoud, 1994; Curran et al., 1992). This is also in line with previous studies that have demonstrated poor signal propagation from leaf to canopy scale (Asner, 1998; Jacquemoud et al., 1996; Yoder and Pettigrew-Crosby, 1995).

According to Combal et al. (2003), three sources of prior information can be distinguished: ancillary data measured on site, knowledge of the type of canopy architecture, and knowledge of the typical distribution of canopy biophysical variables. Combal et al. (2003) as well as Meroni et al. (2004) have shown that utilizing prior information is an efficient way of solving the ill-posed problem and of improving the accuracy of the estimated canopy variables. In the case of spatialized (remote sensing) data, Atzberger (2004) showed that for monospecies canopies the intercorrelation between spectral bands also helps to constrain the ill-posed inverse problem. Extensive field measurement in this study allowed us to identify the maximum and minimum values for the three parameters LAI, LCC and ALA, which increased the sampling density and facilitated the estimation of grass biophysical characteristics.

For several sample plots at least one of the other 6 retrieved parameters (\(C_m, C_a, \text{scale, ALA, hot and N}\) reached the upper/lower boundary. We argue that the possible reason is that some wavelengths are either badly measured or not well modeled by the combined SAILH and PROSPECT canopy reflectance model. Similar results have been found by Schlerf and Atzberger (2006) who demonstrated that the PROSPECT leaf optical properties model is not simulating equally well the leaf optical properties across the 400–2500 nm wavelength range. The possible explanation of too restricted parameter ranges can be excluded as much wider ranges did not ameliorate the results (e.g. LAI: \(R^2=0.37, \text{RMSE}=2.46; \text{CCC: } R^2=0.56, \text{RMSE}=1.27\)).

Selecting subsets of wavelengths derived by Darvishzadeh et al. (in press) from stepwise linear regression gave significantly
higher errors for LAI and leaf chlorophyll content. This was not the case when the selection was based on literature results (subset II) or when only those wavelengths were chosen which were “well” modeled by PROSAIL (subset III). In these cases results similar to those obtained using all wavebands were obtained. The band selection from literature worked well, probably because we only considered wavelengths related to both biophysical and biochemical properties of vegetation, thus maximizing the information content in the input variables while eliminating all other wavelengths that introduce noise and model errors. Similarly, by eliminating wavelength having a high AAE (subset III), we eliminated noisy/badly modeled wavelengths. In the present study, we did not test whether including in the cost function the reflectance uncertainty matrix as shown for example by Lavergne et al. (2007) would improve the results. This would require to run the LUT inversion several times and to use bootstrap techniques to avoid loosing independency between measured and estimated biophysical variables.

It has been demonstrated (Meroni et al., 2004; Schlerf and Atzberger, 2006) that the selection of a few wavebands will give often better results than those achieved using the full spectral resolution. The results in the present study (when using all bands) indicated a relatively good representation of the measured spectra by the PROSAIL model over most spectral regions (see Figs. 3 and 4). Consequently, spectral subsetting did not clearly improve the parameter retrieval. The results of this study confirm that grass canopy characteristics such as leaf area index and canopy chlorophyll content can be estimated through the inversion of a radiative transfer model using hyperspectral measurements with accuracies comparable to those of empirical approaches (Darvishzadeh et al., in press), which is also supported by previous studies (Gemmel et al., 2002; Schlerf and Atzberger, 2006). In contrast to empirical approaches, ground measured biophysical data may be almost entirely used for validating the retrieved model parameters (and to set the LUT ranges) and are not used to calibrate the radiative transfer model (except the soil reflectance spectra which has to be input into the radiative transfer model). Once an appropriate LUT has been built, it can in principle be applied to different remote sensing data acquired over similar vegetation types.

5. Conclusion

This study selected the widely used PROSAIL model to describe the radiation transfer in a heterogeneous Mediterranean grassland for use with hyperspectral data. For fast model inversion, a LUT approach was used. The LUT was built taking into account prior knowledge regarding LAI, LCC and ALA measured in the field. The accuracies of the retrieved vegetation variables are discussed on the basis of (i) the role of different ways of selecting the solution from the LUT (i.e., the best fitting spectra against the mean/median of the best 10 to 100 solutions), (ii) the stratification of data based on species heterogeneity, and (iii) the influence of spectral subsetting. We have demonstrated that the retrieval of canopy chlorophyll content and LAI at canopy level is feasible. However, accuracy decreases if the number of species within a subplot increases. This shows that the selected radiative transfer model is not well adapted to multi-species canopies.

Several authors have used the PROSAIL model in homogeneous crop canopies. Its applicability to heterogeneous grasslands requires further experiments and validation work using different hyperspectral data sets. In this way scale and sensor effects as well as phenological influences can be studied. These factors may lead to (partially) different results.

Unfortunately, the turbid medium assumption used in this model does not account for heterogeneities in the canopies (e.g., clumping effects, multiple leaf layers having different optical characteristics). Therefore, when the turbid medium hypotheses are violated, the model cannot realistically simulate the canopy reflectance, and the retrieved biophysical variables are expected to be biased (Meroni et al., 2004). Improvement in parameter retrieval may be expected from models that explicitly take into account canopy heterogeneities such as vertical leaf color gradient and clumping effects (e.g., Verhoeven and Bach, 2007). However, heterogeneity is a relative term and is strongly scale dependent. Further studies are required to cope with the ill-posed inverse problem when inverting physically based radiative transfer models (e.g., Durbha et al., 2007).

Acknowledgements

We would like to acknowledge the assistance of the park management of Majella National Park, Italy, and in particular of Dr. Teodoro Andrisano. Special thanks go to Prof. Herbert Prins, Dr. Fabio Corsi, Dr. Sip van Wieren, Dr. Moses Cho and Dr. Istiak Sobhan for their assistance during the field campaign. The comments of the two anonymous reviewers have been very helpful in improving the manuscript.

References


Mathworks (2007). Matlab, the language of technical computing. USA: Mathworks, INC.


