Analysis of imaging spectrometer data to evaluate the biochemical content of vegetation, based on the results of a laboratory experiment.

Jean Verdebout, Stéphane Jacquemoud, Giovanni Andreoli, Brian Hosgood,
Alfredo Pedrini and Guido Schmuck

Joint Research Centre, Institute for Remote Sensing Applications
TP 272 - 21020 - Ispra (VA), Italy

ABSTRACT

The reflectance spectrum of green vegetation is mainly determined by the leaf content in chlorophyll and water, the respective spectral signatures being modulated by the structural characteristics of the leaf and the canopy. However, correlation between the reflectance spectra and the leaf content in such constituents as lignin, cellulose, nitrogen, starch, etc. has been demonstrated. To further study this question, we have constituted a data set associating high quality spectra (of single leaves, optically thick stacks, needles, stalks, on fresh and dried material) with a number of physical and chemical measurements (leaf thickness, water content, chlorophyll, carotenoids, cellulose, lignin, proteins, nitrogen, starch). This data set has been used to investigate the link between the optical properties and the composition, focusing on the biochemical components. Two approaches have been followed: the classical regression analysis and modelling based on the Kubelka-Munk formula. In the latter case, empirical specific absorption coefficients have been determined and then used to decompose the infrared spectra in water and other components contributions. This method is successful in retrieving the relative water content but does not yet allow to estimate the biochemical content. It was applied both on laboratory and AVIRIS spectra.

2. INTRODUCTION

Estimating leaf biochemistry and leaf water status with remote sensing data is a challenge for the years to come. In the framework of the Global Change Program, it should provide key information about the functioning of terrestrial ecosystems by extending ecological models to different scales[1]. It has also implications in agriculture to follow crop development and yield predictions. The biochemical constituents of interest are lignin, proteins (nitrogen), cellulose and starch, as well as chlorophyll and foliar water[2]. The major processes involved in the terrestrial ecosystem like photosynthesis, primary production, or litter decomposition can be related to these constituents. For example, according to Wessman[3], the lignin / nitrogen ratio drives the decomposition rates of forest litter and is an indicator of ecosystem processes. As leaves are the most important surfaces of a plant canopy, relating their optical properties to these constituents is a priority. Two different approaches may be considered: the first one rests on a theoretical basis which consists in developing a leaf scattering and absorption model involving biochemistry. For the moment, only chlorophyll content and water content have been explicitly included in leaf optical properties models. Conel et al.[4] recently proposed a two-stream radiative transfer model to analyse the influence of protein, cellulose, lignin and starch on leaf reflectance, but this model has not been validated. Except for this work, only statistical analyses have been performed to retrieve leaf biochemical components. They represent the second approach.

Historically, statistical methods were used by people involved in forage quality analyses. They developed near infrared reflectance spectroscopy (NIRS), a rapid spectral analysis technique for determination of vegetation biochemical contents. The application of NIRS for remote sensing purpose is rather recent. It was first used on dried, ground material. Multiple stepwise regressions were mainly performed using dry leaves or needles, when they did not ensue directly from vegetation powders[5,6,7]. Surprising as it may seem, only a few studies deal with fresh and green leaves[8,9]. The situation is more complex regarding fresh material because of dominant water absorption beyond 1.0 μm and different scattering phase function. Calibration equations derived from dried material cannot be applied directly to fresh leaves and new wavelengths have to be selected.

We intend in that paper to investigate the retrieval of chlorophylls, water, protein, cellulose, lignin, and starch by statistics both on fresh and dry material, on individual leaves and on optically thick samples (stacked leaves + needles or powders).

3. THE EXPERIMENT

Measurements from a laboratory experiment organised in the Joint Research Centre during the summer of 1993 were used. We have undertaken to build a unique data set associating visible/infrared spectra of vegetation elements (leaves, conifer needles,

stems, etc.) both with physical measurements and biochemical analyses. In order to have a wide range of variation of leaf internal structure, pigmentation, water content, and biochemical components, plant species with different types of leaves have been collected outdoors after a space of two months. About 70 leaf samples representing 50 species of woody and herbaceous plants were obtained from trees and crops in the area of the JRC (Table 1). In addition, various substances such as powdered starch or proteins and vegetation material such as stems or bark were also included in the data set.

Gymnosperms	Monocotyledons	Dicotyledons
Picea abies, Pinus contorta, Pinus wallichiana, Pseudotsuga menziesii	*Bambusa acundinacea, Chamaerops humilis, Iris germanica L., Musa ensete, Oryza sativa, Phleum pratense L., Phragmites communis, Sorghum halepense, Zea Mays L.*	*Acer Pseudo-Platanus L., Alnus glutinosa, Armeniaca vulgaris, Beta vulgaris L., Betula alba L., Brassica oleracea L., Castanea sativa, Corylus avellana L., Fagus silvatica L., Ficus carica L., Fraxinus excelsior L., Hedera helix L., Helianthus annuus L., Juglans regia L., Lactuca sativa, Laurus nobilis L., Lycopersicum esculentum, Medicago sativa L., Morus alba L., Platanus acerifolia, Populus canadensis, Populus tremula L., Prunus serotina, Prunus laurocerasus, Quercus pubescent, Quercus rubra, Robinia Pseudacacia L., Salix alba L., Salvia officinalis L., Soja hispida, Solanum tuberosum L., Tilia platyphylla, Trifolium pratense L., Ulmus glabra, Urtica dioica L., Vitis silvestris, Vitis vinifera L.*

Table 1. Scientific names of plant leaves used in this study.

stems, etc.) both with physical measurements and biochemical analyses. In order to have a wide range of variation of leaf internal structure, pigmentation, water content, and biochemical components, plant species with different types of leaves have been collected outdoors after a space of two months. About 70 leaf samples representing 50 species of woody and herbaceous plants were obtained from trees and crops in the area of the JRC (Table 1). In addition, various substances such as powdered starch or proteins and vegetation material such as stems or bark were also included in the data set.

A Perkin Elmer Lambda 19 spectrophotometer equipped with an integrating sphere allowed the measurement of the directional-hemispherical reflectance (R) and transmittance (T) of the upper faces of leaves. The absorptance (A) was derived from R and T through the simple relationship: $A=1-(R+T)$. Moreover, the reflectance of optically thick samples ($R\infty$) was obtained with needles and by stacking leaves in order to magnify the radiometric signal and minimise the leaf to leaf variability. Spectra were scanned over the 400-2500 nm wavelength interval with 1 nm step. The spectral resolution varied from 1 to 2 nm in the visible / near infrared (400-1000 nm) and from 4 to 5 nm in the middle infrared (1000-2500 nm). The calibration of the instrument was performed using Spectralon reflectance and wavelength calibration standards. For each sample, the optical properties of 5 representative fresh leaves were averaged to reduce the noise of the spectrophotometer and to smooth out the small but not negligible leaf to leaf variability. All the above procedure was repeated on dried leaves and needles to analyse the influence of water which is known to obscure the biochemical information in the middle infrared region. The wavelength interval was increased from 1 nm to 5 nm in order to reduce the noise, the number of data, and consequently the calculation time. In total, 421 narrow wavebands were available. The first derivative was calculated for each spectrum: noise in spectra was low enough to avoid the use the smoothing techniques. Table 3 summarises the radiometric data acquired in the framework of this experiment.

Parallel to the radiometric acquisitions, many physical and biological measurements were performed on leaf samples. Leaf blade thickness was measured with a calliper rule (5 measurements per leaf). We immediately measured the fresh weight of a 4.10 cm^2 disc taken on each leaf using a cork borer. Then, the disc was placed in a drying oven at 85°C for 48 hours and re-weighed to determine the relative water content (RWC = water mass over fresh mass), the equivalent water thickness or water depth (EWT = water mass per unit leaf area), and the specific leaf area (SLA = dry weight per unit leaf area). Special attention has been paid to the measurement of SLA because biochemical concentrations used in leaf optical properties models are generally expressed in weights per unit leaf area; in this way, we will be able in the future to compare outputs from models with outputs from statistical relationships. Remaining leaf samples have been frozen for later biochemical analysis: the photometric determination of photosynthetic pigments (chlorophyll a, b, and total carotenoids) was performed with a UV-2001 PC spectrophotometer in 100% acetone using the equations of Lichtenthaler[10]. With regard to the other biochemical constituents, about 250 g of fresh material were placed in a drying oven and were sent to two independent laboratories (Lab I and Lab II) which were in charge of the measurements of total proteins, cellulose, lignin, and starch using standard wet chemical analyses. A summary of the results is given in Table 2.

	Lab. I			Lab. II		
	range	mean	std	range	mean	std
Protein	6.44-36.76	18.41	7.66	6.06-35.75	18.75	8.05
Cellulose	9.06-37.17	21.03	6.97	2.10-37.50	17.49	6.83
Lignin	1.09-27.48	10.77	6.17	0.03-23.15	8.06	5.13
Starch	0-9.99	1.78	2.05	0-9.42	1.35	2.07

Table 2. Summary of the biochemical analysis, values are expressed in percent of dry weight.

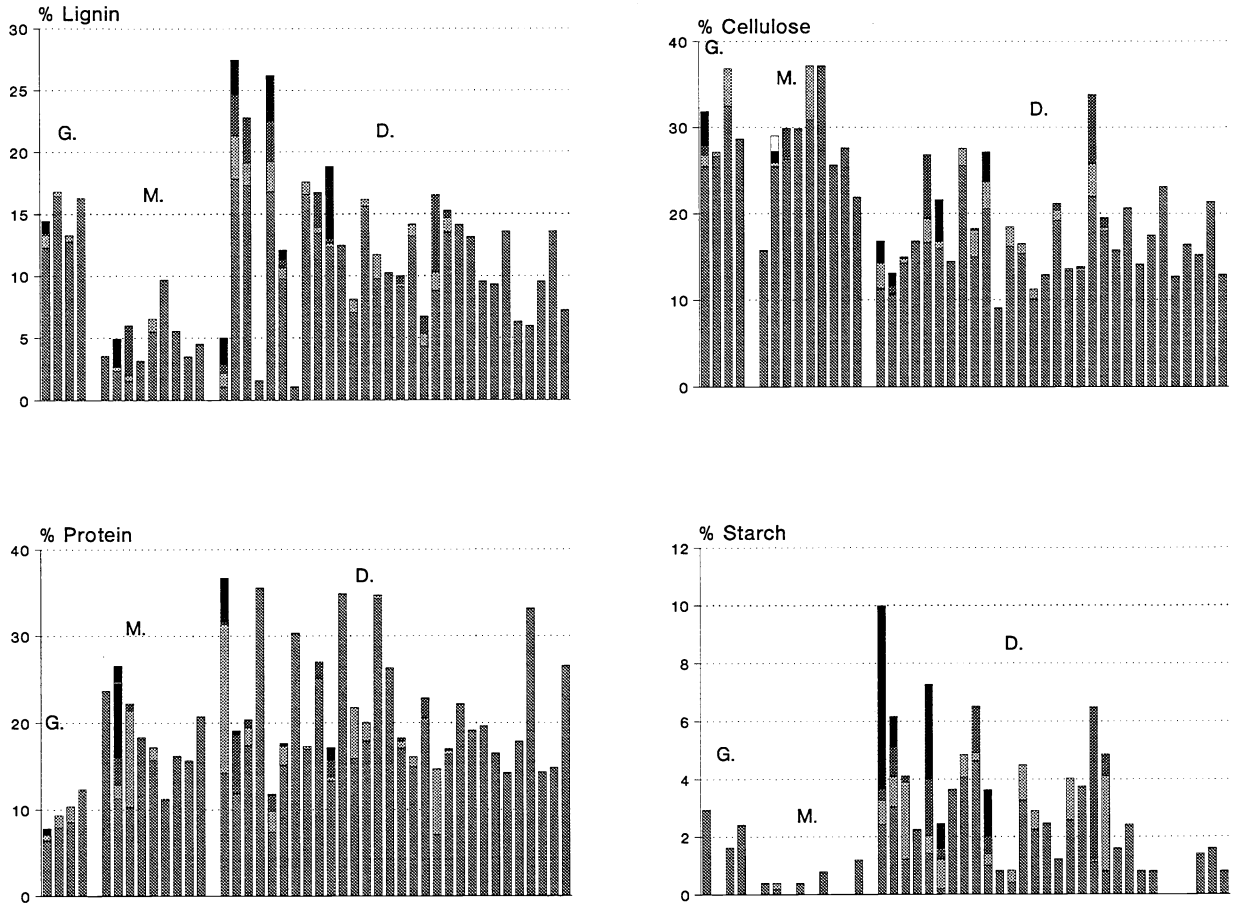


Figure 1. Distributions of lignin, cellulose, protein, and starch concentrations for Gymnosperms (G), Monocotyledons (M), and Dicotyledons (D).

The comparison between the concentration values (g/g) provided by the two laboratories gave us an idea of the precision of these analyses: protein and cellulose measurements were quite consistent while lignin and starch measurements differed significantly. As explained by Curran[11], these discrepancies may be imputed to the different methods of chemical extraction. The chemistry data spanned a broad range of values (Table 2). In the following, only results of Lab I are presented. Figure 1 shows the distribution of chemistry data for 10 Gymnosperm, 15 Monocotyledon, and 48 Dicotyledon samples representing 50 species: the lignin concentration of needles and Dicotyledon leaves which mainly correspond to woody plants respectively amounts to about 14.3% and 12.1% of the dry weight while that of Monocotyledon leaves is three times lower (4.3%). The

distribution is rather different for cellulose: its concentration in needles and Monocotyledons respectively amounts to 29.5% and 28.3% while that in Dicotyledons is 10% lower. Lignin and cellulose are structural components of cell walls: they constitute all together around 30.8% of the dry weight of plant leaves and 43.8% of the dry weight of needles, but they are physiologically and biochemically inactive[12]. In contrast, the concentration of protein which has essential roles in biochemical and physiological processes is twice higher in plant leaves (19.4%) than in needles (8.5%). The C/N ratio which indicates changes in decomposition rates affecting nutrient cycling and trace gas fluxes[3] consequently varies from 1 to 3. Finally, the starch concentration is very small except for Dicotyledons where it equals 2.5% of the dry weight. The correlation between the chemical concentrations are rather low (Figure 2).

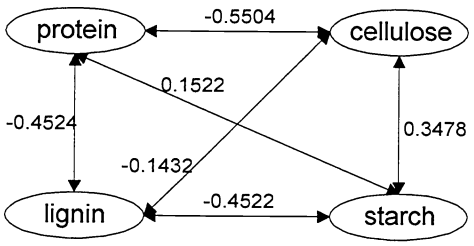


Figure 2. Correlation coefficient for relationships among chemicals on a dry weight bases (g/g).

4. THE REGRESSION ANALYSIS

For determining predictive spectral wavebands of chemical concentrations, stepwise multiple regression analysis which consists in relating the concentration C with the reflectance (or other spectral property) at 1, 2,...,N wavelengths [$R(\lambda 1)$, $R(\lambda 2)$,...,$R(\lambda N)$] were performed with a maximum of ten regressors:

$$C = a0 + a1.R(\lambda 1) + a2.R(\lambda 2) + ... + aN.R(\lambda N) \quad (1)$$

The regression model was constructed adding the independent variables (reflectances for instance) one at a time. The first step was to choose the single variable which is the best statistical predictor; the second independent variable to be added to the regression equation was that which provides the best fit in conjunction with the first variable. Further variables were then added in this recursive fashion, adding at each step the optimum variable, given the other variables already in the equation. Goodness of fit was measured by two criteria: the squared multiple correlation coefficient R^2 which can be interpreted as the fraction of total sum of squares explained by the regression, and by the root mean square error *rmse* calculated as

$$rmse = \sqrt{\frac{\sum (C - C^*)}{n}} \quad (2)$$

where C and C* are respectively the measured and estimated concentrations, and n is the number of measurements.

5. RESULTS AND DISCUSSION

Each chemical component (protein, cellulose, lignin, starch) was treated independently of the others, and so each yields an independent set of regression coefficients. We distinguished two groups of plant material: individual leaves (R, T, and A) and stacked individual leaves + needles ($R\infty$) which represented optically thick samples. First derivatives – respectively $\partial R/\partial\lambda$, $\partial T/\partial\lambda$, $\partial A/\partial\lambda$, and $\partial R\infty/\partial\lambda$ – and $\log(1/R\infty)$ were also investigated. In each group, fresh leaves were separated from dry leaves. Tables 3 and 4 detail statistical outputs of the stepwise multiple regression analysis for proteins, cellulose, lignin, and starch.

From a general point of view, correlations are better for optically thick samples than for individual leaves, and better for dry samples than for fresh samples. Best R^2 and rmse values are obtained for first derivatives. As for individual leaves, $\partial R/\partial\lambda$, $\partial T/\partial\lambda$, and $\partial A/\partial\lambda$ provide very similar results. As for optically thick samples, $\log(1/R\infty)$ which can be considered as an approximation of absorptance does not lead to significant improvement of correlations and, as noticed earlier by Gastellu-Etchegorry et al.[7], it is surprisingly a worse predictor of leaf biochemistry than the raw infinite reflectance. In general, total proteins, cellulose, and lignin are reasonably estimated with R^2 higher than 0.70 with 5 wavelengths, and R^2 higher than 0.85 with 10 wavelengths. Starch whose content in the leaves is very small (<3% of dry matter) is retrieved with more difficulty.

Let us detail these results with the case of total proteins. Figure 3 shows the variation of R^2 as terms are added to the regression equation; curves are very similar to that obtained by Card et al.[5]. One can note a strong increase of R^2 for the first five wavelengths and a saturation effect when the number of wavelengths approaches ten. The other constituents present the same trends (results not shown).

The distribution of the wavelengths stemming from the stepwise multiple regression analysis is quite amazing. To illustrate it, the reflectance and transmittance spectra both of a typical fresh and dry leaf have been plotted together with the position of the first, second, third, until tenth wavelength selected in the estimation of proteins (Figures 4 and 5). First, one can notice great differences from one case to another: in particular for fresh leaves, spectral regions sensitive to the protein concentration are completely different. One can also notice groups of wavelengths very close together. According to Card et al. [5], it is difficult to associate particular chemical bonds with the wavelengths selected by stepwise regression since the latter depend on many factors such as the kind of data chosen (reflectance, transmittance, or absorptance). Moreover, leaves contain several constituents each with a number of absorption peaks, causing peak broadening and shifting, and so selected wavelengths do not always occur precisely at known stretching frequencies[11]: by simulating the effects on spectral reflectance of hypothetical mixtures of protein, cellulose, lignin and starch, Conel et al.[4] underscored such effects which complicate the interpretation of leaf optical properties in terms of leaf biochemistry.

Nevertheless, Figure 6 shows potential for predicting chemical concentrations over a wide range of conditions: the plots all indicate a good distribution for the biochemical components and a low bias in the regression as shown by symmetry of points about the 45° line through the origin.

			Proteins						Lignin					
			N=1		N=5		N=10		N=1		N=5		N=10	
			R^2	rmse	R^2	rmse	R^2	rmse	R^2	rmse	R^2	rmse	R^2	rmse
individual leaves	fresh (63)	R	0.1787	6.30	0.4998	4.91	0.8030	3.08	0.1870	5.77	0.4403	4.79	0.6847	3.59
		$\partial R/\partial\lambda$	0.3915	5.42	0.7074	3.76	0.8634	2.57	0.3023	5.34	0.6386	3.85	0.8200	2.71
		T	0.0897	6.63	0.5557	4.63	0.8132	3.00	0.0784	6.14	0.4946	4.55	0.6630	3.71
		$\partial T/\partial\lambda$	0.4148	5.32	0.7806	3.25	0.8836	2.37	0.3094	5.32	0.6087	4.00	0.7808	3.00
		A	0.2212	6.13	0.4976	4.93	0.8267	2.89	0.2044	5.71	0.3980	4.96	0.6089	4.00
		$\partial A/\partial\lambda$	0.4096	5.34	0.7775	3.28	0.9008	2.19	0.3475	5.15	0.6800	3.62	0.8492	2.49
	dry (57)	R	0.3636	5.50	0.8100	3.01	0.8935	2.25	0.1865	5.66	0.4812	4.52	0.6089	3.92
		$\partial R/\partial\lambda$	0.7489	3.46	0.8745	2.45	0.9400	1.69	0.3943	4.88	0.7405	3.20	0.8500	2.43
		T	0.2974	5.78	0.7271	3.61	0.8623	2.56	0.1003	5.95	0.6838	3.53	0.7974	2.82
		$\partial T/\partial\lambda$	0.7605	3.38	0.8870	2.32	0.9468	1.59	0.4092	4.82	0.7161	3.34	0.8497	2.43
		A	0.4632	5.06	0.7517	3.44	0.9050	2.13	0.1751	5.70	0.5605	4.16	0.7477	3.15
		$\partial A/\partial\lambda$	0.7337	3.56	0.8744	2.45	0.9454	1.61	0.4102	4.82	0.7220	3.31	0.8360	2.54
stacked leaves + needles	fresh (73)	$R\infty$	0.4790	5.49	0.7420	3.86	0.8432	3.01	0.1402	5.69	0.3921	4.78	0.5700	4.02
		$\partial R\infty/\partial\lambda$	0.6159	4.71	0.8011	3.39	0.9099	2.28	0.2606	5.28	0.5897	3.93	0.7526	3.05
		$\log(1/R\infty)$	0.4528	5.63	0.7556	3.76	0.8518	2.93	0.1482	5.66	0.3782	4.84	0.5598	4.07
	dry (67)	$R\infty$	0.5860	4.83	0.8568	2.84	0.8977	2.40	0.1535	5.50	0.4294	4.51	0.6873	3.34
		$\partial R\infty/\partial\lambda$	0.7868	3.46	0.8884	2.51	0.9420	1.81	0.4031	4.62	0.7225	3.15	0.8708	2.15
		$\log(1/R\infty)$	0.5325	5.13	0.7926	3.41	0.8752	2.65	0.1235	5.60	0.4044	4.61	0.6450	3.56

Table 3. Results of stepwise multiple regression analysis on protein and lignin for N=1, 5, 10 regressors. The number of samples used to estimate the chemical concentrations are shown in brackets.

			Cellulose						Starch					
			N=1		N=5		N=10		N=1		N=5		N=10	
			R^2	rmse	R^2	rmse	R^2	rmse	R^2	rmse	R^2	rmse	R^2	rmse
individual leaves	fresh (63)	R	0.0399	6.25	0.4948	4.53	0.6159	3.95	0.0461	2.05	0.3587	1.68	0.6514	1.24
		∂R/∂λ	0.2366	5.57	0.6160	3.95	0.7750	3.03	0.1557	1.93	0.5367	1.43	0.7387	1.07
		T	0.0571	6.20	0.3497	5.15	0.6503	3.77	0.0650	2.03	0.3435	1.70	0.6005	1.33
		∂T/∂λ	0.2867	5.39	0.6284	3.89	0.8371	2.58	0.2838	1.78	0.5551	1.40	0.7075	1.14
		A	0.0821	6.11	0.4592	4.69	0.6584	3.73	0.0473	2.05	0.3292	1.72	0.5788	1.36
		∂A/∂λ	0.4325	4.81	0.7147	3.41	0.8443	2.52	0.2330	1.84	0.4998	1.49	0.7091	1.13
	dry (57)	R	0.1371	5.87	0.6055	3.97	0.8366	2.54	0.0401	1.87	0.4077	1.47	0.5940	1.22
		∂R/∂λ	0.4095	4.86	0.8198	2.68	0.9282	1.69	0.2397	1.67	0.6389	1.15	0.8537	0.73
		T	0.0369	6.20	0.6861	3.54	0.8902	2.09	0.0419	1.87	0.5749	1.25	0.7082	1.03
		∂T/∂λ	0.3877	4.94	0.8435	2.50	0.9237	1.75	0.2648	1.64	0.6839	1.07	0.8442	0.75
		A	0.1301	5.89	0.5388	4.29	0.8616	2.35	0.0461	1.87	0.3588	1.53	0.5807	1.24
		∂A/∂λ	0.4015	4.89	0.8155	2.71	0.9248	1.73	0.2478	1.66	0.7219	1.01	0.8473	0.75
stacked leaves + needles	fresh (73)	R∞	0.2480	6.00	0.5960	4.40	0.7047	3.76	0.1289	1.90	0.3571	1.64	0.4490	1.51
		∂R∞/∂λ	0.4188	5.28	0.6986	3.80	0.8525	2.66	0.2293	1.79	0.4775	1.47	0.6698	1.17
		log(1/R∞)	0.2436	6.02	0.5976	4.39	0.7452	3.49	0.1276	1.90	0.4011	1.58	0.5167	1.42
	dry (67)	R∞	0.3522	5.57	0.6142	4.30	0.6869	3.87	0.0795	1.79	0.3730	1.47	0.6085	1.17
		∂R∞/∂λ	0.4739	5.02	0.8118	3.00	0.9069	2.11	0.2462	1.62	0.6397	1.12	0.7927	0.85
		log(1/R∞)	0.3631	5.52	0.5437	4.68	0.8351	2.81	0.0897	1.78	0.4432	1.39	0.5962	1.18

Table 4. Results of stepwise multiple regression analysis on cellulose and starch for N=1, 5, 10 regressors. The number of samples used to estimate the chemical concentrations are shown in brackets.

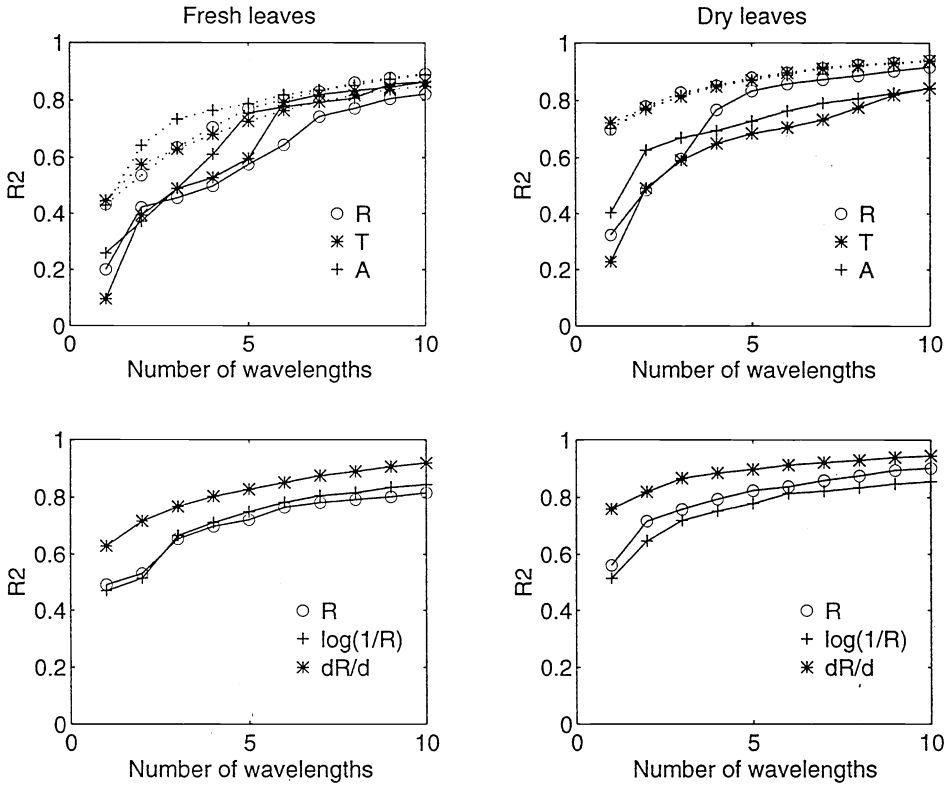


Figure 3. Multiple squared correlation coefficient R^2 versus the number of terms in the regression equation for the stepwise regression of protein concentration. Plots at the top represent the individual leaves [o reflectance, * transmittance, + absorptance, — raw values, ··· first derivative], plots at the bottom the optically thick samples (stacked leaves + needles) [infinite reflectance, o raw values, + first derivative, * log(1/R)].

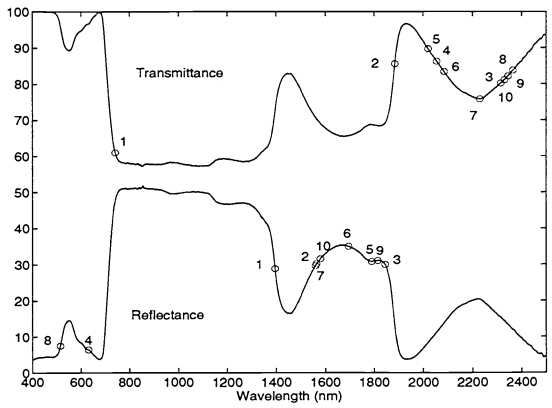


Figure 4. Reflectance and transmittance spectra of a typical fresh leaf. Wavelengths selected by the stepwise multiple regression analysis to explain protein are represented by circles.

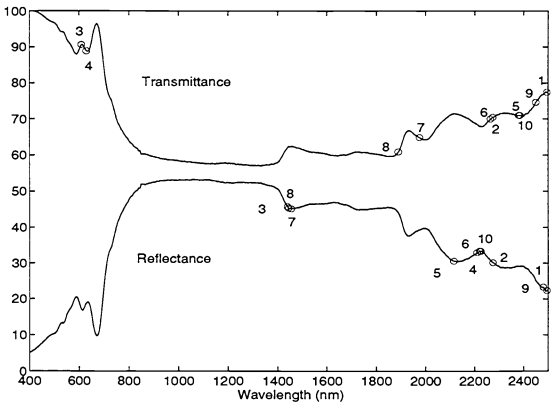


Figure 5. Reflectance and transmittance spectra of a typical dry leaf. Wavelengths selected by the stepwise multiple regression analysis to explain protein are represented by circles.

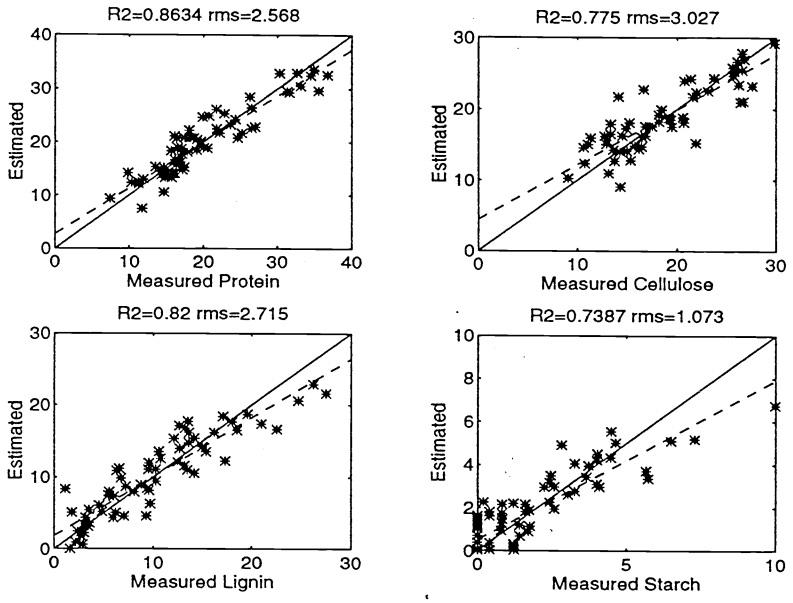


Figure 6. Protein, cellulose, lignin, and starch predicted concentrations (fresh individual leaves, $\partial R/\partial\lambda$, 10 wavelengths) versus concentrations according to wet chemical analyses.

6. USE OF KUBELKA-MUNK FORMULA TO DETERMINE EMPIRICAL ABSORPTION COEFFICIENTS

A more promising approach in the longer term is to include the biochemical components in radiative transfer models of the vegetation spectrum (such as the PROSPECT model[13]). However, to do this, one needs the corresponding specific absorption coefficients and there are not available. There exist spectrometric measurements performed on "pure substances" (cotton for cellulose, starch powder, etc..). These can indicate where the absorption features are expected to be found but do not provide quantitative values for the absorption coefficient, the reflectance spectra being also determined by the unknown scattering

characteristics of the samples. Furthermore, for some components (in particular lignin) it is actually very difficult if not impossible to produce pure samples. In the following, we present a methodology to deduce empirical "in vivo" absorption coefficients from the reflectance spectra of fresh leaves stacks.

A first assumption is to assimilate the reflectance of a leaves stack to the diffuse reflectance of a semi-infinite homogeneous medium. This is a strong assumption which ignores the specular component (known to be of the order of a few percent) and the layered structure of leaves . Regarding the second point, the assumption is probably not valid at wavelengths for which the absorption is highest , as in the absorption bands of water at 1.45 and 1.92 μm. At 1.92 μm, the absorption coefficient of water reaches 118 cm^{-1}, the corresponding penetration length is 85 μm and is less than a typical leaf thickness. In this case, the reflectance is no longer determined by the whole leaf tissue but by its first layer which may have different characteristics. For this reason and also because these bands are accessible to imaging spectrometry , we will restrict the analysis to two spectral ranges: from 1.5 to 1.8 μm and 2.1 to 2.4 μm.

The Kubelka-Munk formula for a semi -infinite medium relates the reflectance to the single scattering albedo and thereby to the absorption and scattering coefficients of the medium:

$$R(\lambda) = \frac{2 - \omega_0(\lambda) - 2.\sqrt{1-\omega_0(\lambda)}}{\omega_0(\lambda)} \tag{3}$$

$$\text{with} \quad \omega_0 = \frac{s(\lambda)}{s(\lambda)+k(\lambda)} = \frac{1}{1+\frac{k(\lambda)}{s(\lambda)}} \tag{4}$$

where λ is the wavelength

$R(\lambda)$ is the reflectance

$\omega_0(\lambda)$ is the single scattering albedo of the medium

$s(\lambda)$ is the scattering coefficient of the medium

$k(\lambda)$ is the absorption coefficient of the medium

The Kubelka-Munk formula can be inverted to obtain the ratio of the absorption to scattering coefficients:

$$\frac{k(\lambda)}{s(\lambda)} = \frac{(R(\lambda)+1)^2}{4.R(\lambda)} - 1 = KM^{-1}(R(\lambda)) \tag{5}$$

Considering the leaf tissue as a homogeneous mixture, the absorption coefficient is then expressed as a linear combination of the specific absorption coefficients of the components:

$$k(\lambda) = rwc.k_w(\lambda) + (1-rwc).\sum_{i=1}^{N_c} c_i.k_i(\lambda) \tag{6}$$

$$\text{with} \quad \sum_{i=1}^{N_c} c_i = 1 \tag{7}$$

where rwc is the leaf relative water content (fresh weight-dry weight)/fresh weight

$k_w(\lambda)$ is the specific absorption coefficient of water , the measurement by Curcio and Petty [14] was used

c_i is the concentration of component i (fraction of the leaf dry weight)

$k_i(\lambda)$ is the specific absorption coefficient of component i

N_c is the number of components

Note: imposing that the weight fractions of the components sum to one implies, in practice, that the last component is not better identified than being the residual , with a weight fraction equal to one minus the sum of the others.

The scattering coefficient depends on the leaf internal structure and will, a priori, be different for each leaf . As the scattering is mainly due to multiple reflection and refraction at the cell boundaries, and as the refractive index does not depend strongly of wavelength, the scattering coefficient is also expected to show a weak wavelength dependence. However, as the procedure allows it, we will describe $1/s(\lambda)$ with a polynomial of wavelength:

$$\frac{1}{s(\lambda)} = q_0 + q_1.\lambda + ... + q_n.\lambda^n = P_s^n(\lambda) \tag{8}$$

By using equations (5) to (8), we are able to write a system of equations to determine the specific absorption coefficients $k_i(\lambda)$. For each of the considered wavelengths λ and each measured leaf stack m, we have:

$$KM^{-1}(R_m(\lambda)) = (q_o^m + q_1^m.\lambda + ... + q_n^m.\lambda^n).\left[rwc_m.k_w(\lambda) + (1 - rwc_m).\sum_{i=1}^{N_c} c_i^m.k_i(\lambda)\right] \tag{9}$$

In this equation, the relative water content and the components concentrations are known from the biochemical analysis. The unknowns are the polynomial coefficients q_j^m and the absorption coefficients $k_i(\lambda)$, their number is $(n+1).N_m + N_c.N_\lambda$ where N_λ is the number of considered wavelengths. The number of equations is $N_m.N_\lambda$, in general there are many more equations than unknowns and the system is solved in terms of minimising the error. This was performed by using a non linear least mean squares fitting procedure also constraining the absorption coefficients to be positive.

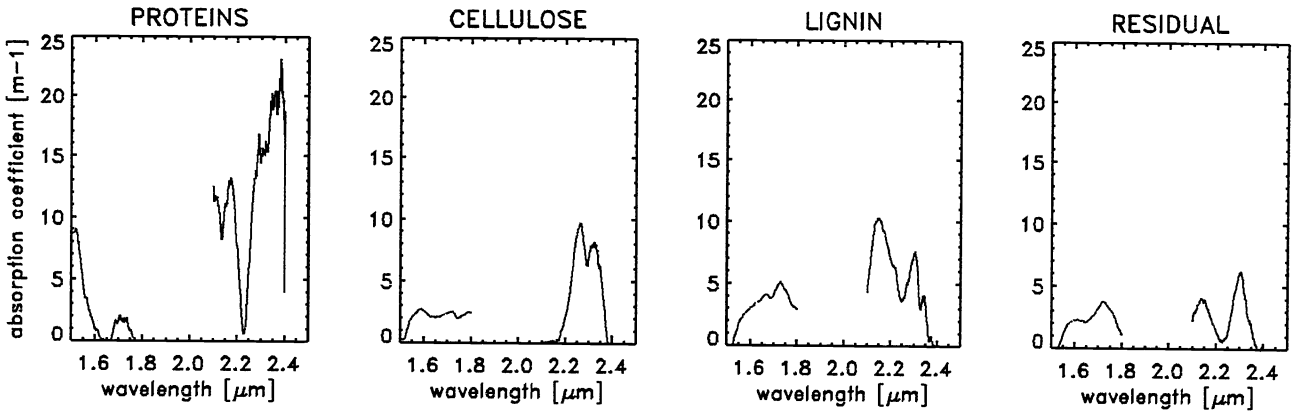


Figure 7. Empirical in vivo absorption coefficients retrieved from the spectra of fresh leaves stacks.

Figure 7. shows the result obtained for the absorption coefficients of proteins, cellulose, lignin and the residual. In this case, 602 wavelengths were considered , 76 measurements were used and the inverse of the scattering coefficient was taken linearly dependent of wavelength (n=1) ; this lead to a system of 45,752 equations for 2,560 unknowns. Although the absorption features do not resemble those observed in the measurements on "pure substances", they were repeatedly found using different degrees of the polynomial and different subsets of measurements. The question should be further examined by using different sets of biochemical components for the decomposition and trying with the dried material. The solution illustrated here is the one providing the best reconstruction of the spectra (taking n>1 no longer improves it significantly).

Figure 8. shows two examples of the decomposition of a spectrum first in its water and dry matter contributions and second the decomposition of the dry matter in proteins, cellulose, lignin and residual contributions. We may first notice that the reconstructed spectrum is almost indistinguishable from the measurement : the average relative difference between the two was 1,8 % (average on wavelength and the 76 samples). The dry matter contribution is small with respect to the water

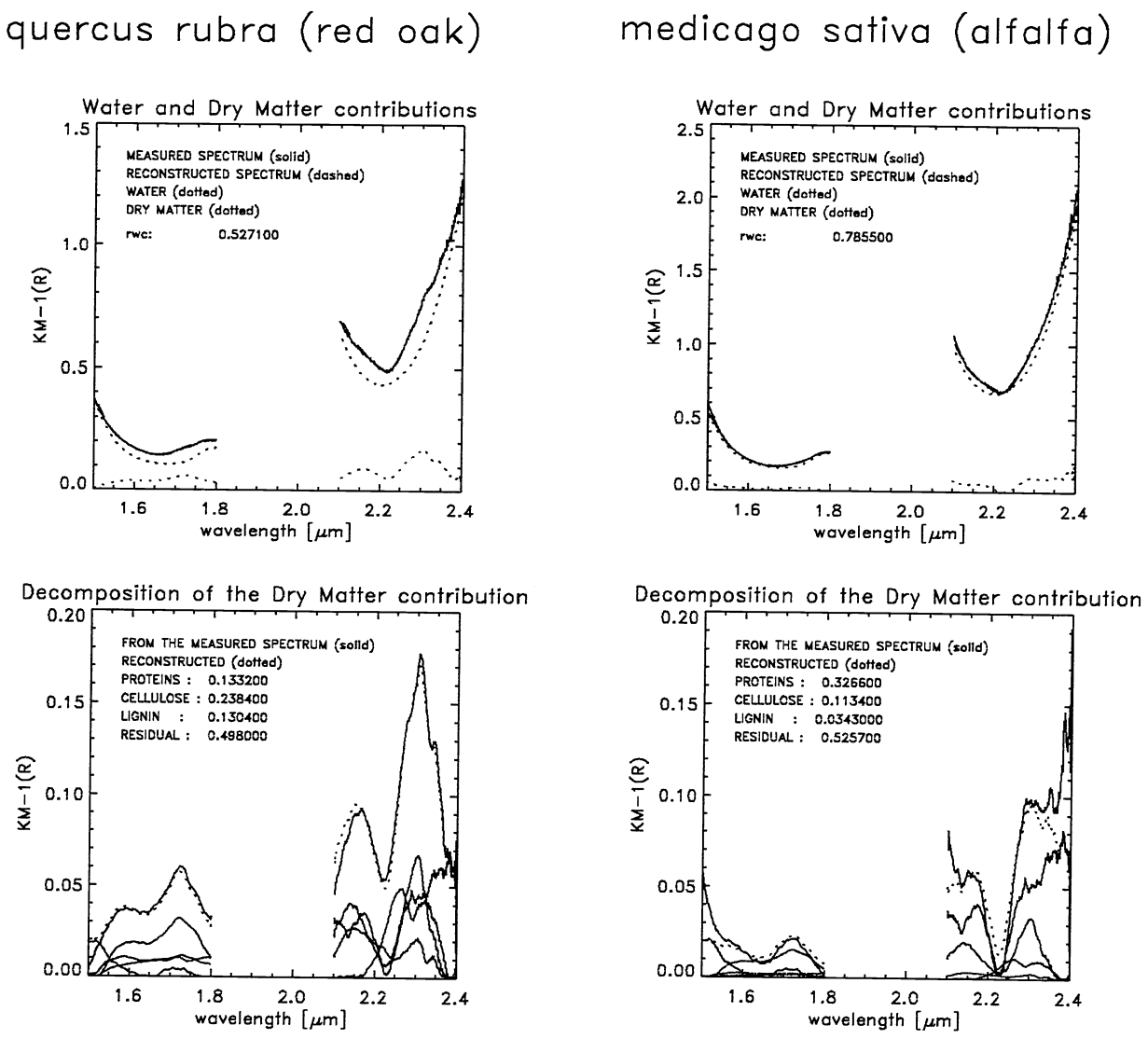


Figure 8. Examples of the decomposition of a spectrum in contributions of water and dry fraction (top) and of the dry fraction in proteins, cellulose, lignin and residual contributions.

contribution and appears to be an intricate combination of the biochemical components contributions of which none is dominant.. This is not really surprising: i. the absorption features are fundamentally associated with chemical bonds ii. cellulose, lignin and proteins (to a lesser extend because they contain nitrogen) have a similar chemical composition and comparable concentrations.

7. INVERSION ON LABORATORY SPECTRA

The purpose of determining absorption coefficients is of course to use them for the inverse problem consisting in evaluating the water and biochemical content of a leaves from their reflectance spectrum. Given a leaves stack reflectance $R(\lambda)$, we fit it according to the same model as before, with n=1, this writes:

$$KM^{-1}(R(\lambda)) = \frac{(R(\lambda)+1)^2}{4.R(\lambda)} - 1 = (q_0 + q_1.\lambda).\left[rwc.k_w(\lambda) + (1-rwc).\sum_{i=1}^{N_c} c_i.k_i(\lambda) \right] \tag{10}$$

$$\text{with } \sum_{i=1}^{N_c} c_i = 1 \tag{11}$$

where q_0, q_1, rwc and the four c_i are the free parameters in the least mean squares fitting procedure (here too the procedure constrains these parameters to be positive).

The inversion was performed on the 76 spectra used in the forward procedure to determine absorption coefficients. Figure 9. shows that the inversion allows a reasonably good estimation of the leaf relative water content. On the contrary, none of the other components concentration could be retrieved in this way. This is not particularly astonishing in view of their complex contribution to the spectrum. We are facing here a common problem with non linear fitting procedures: the existence of several minima in the merit function. As a result, though the spectrum may be very well reconstructed, the set of retrieved concentrations does not correspond to the real one.

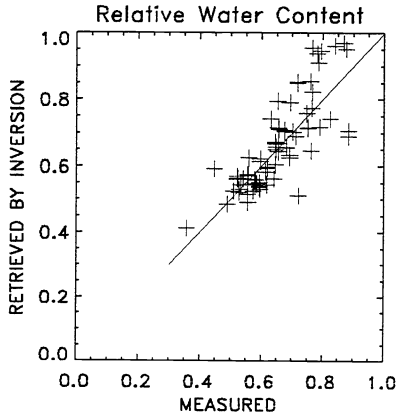


Figure 9. The relative water content retrieved by inversion of the model versus the measured values for 76 spectra of leaves stacks.

8. INVERSION ON IMAGING SPECTROMETER DATA

Without pretending that the reflectance spectrum of a canopy is identical to that of a stack of leaves, the inversion of the model described above on spectra produced by an imaging spectrometer can be meaningful. If we consider canopies such as forests,

the light essentially interacts with the leaves and the effects of multiple scattering (from leaf to leaf) on the spectrum should not be very different from those seen in the stack (the distance between leaves does not play any role). The local shape of the spectrum could be modified by the contribution of the branches, trunks and soil but this will be a minor contribution if the LAI is high enough. The apparent canopy reflectance is influenced by its directional properties and is much lower than the hemispherical reflectance of a stack. Considering equations (10) and (11), one can see that the separation of water and dry fraction will not be influenced by a change of scale in the reflectance, the only effect will be on the magnitude of the polynomial coefficients describing the scattering. In other terms, the mathematical formulation is still meaningful but the scattering coefficient has now to be interpreted as an apparent canopy scattering coefficient which will be influenced both by the leaf internal structure and the canopy structure.

We performed the inversion of the model on a 1992 AVIRIS reflectance scene of Blackhawk Island. The atmospheric correction was done by Roger Clark and Gregg Swayze of USGS. Blackhawk Island is a forested scientific reserve on the Wisconsin river and is a classical test site for studies of canopy biochemistry[15,16]. The forest composition varies along the east-west direction (species present are pine, red oak and sugar maple).

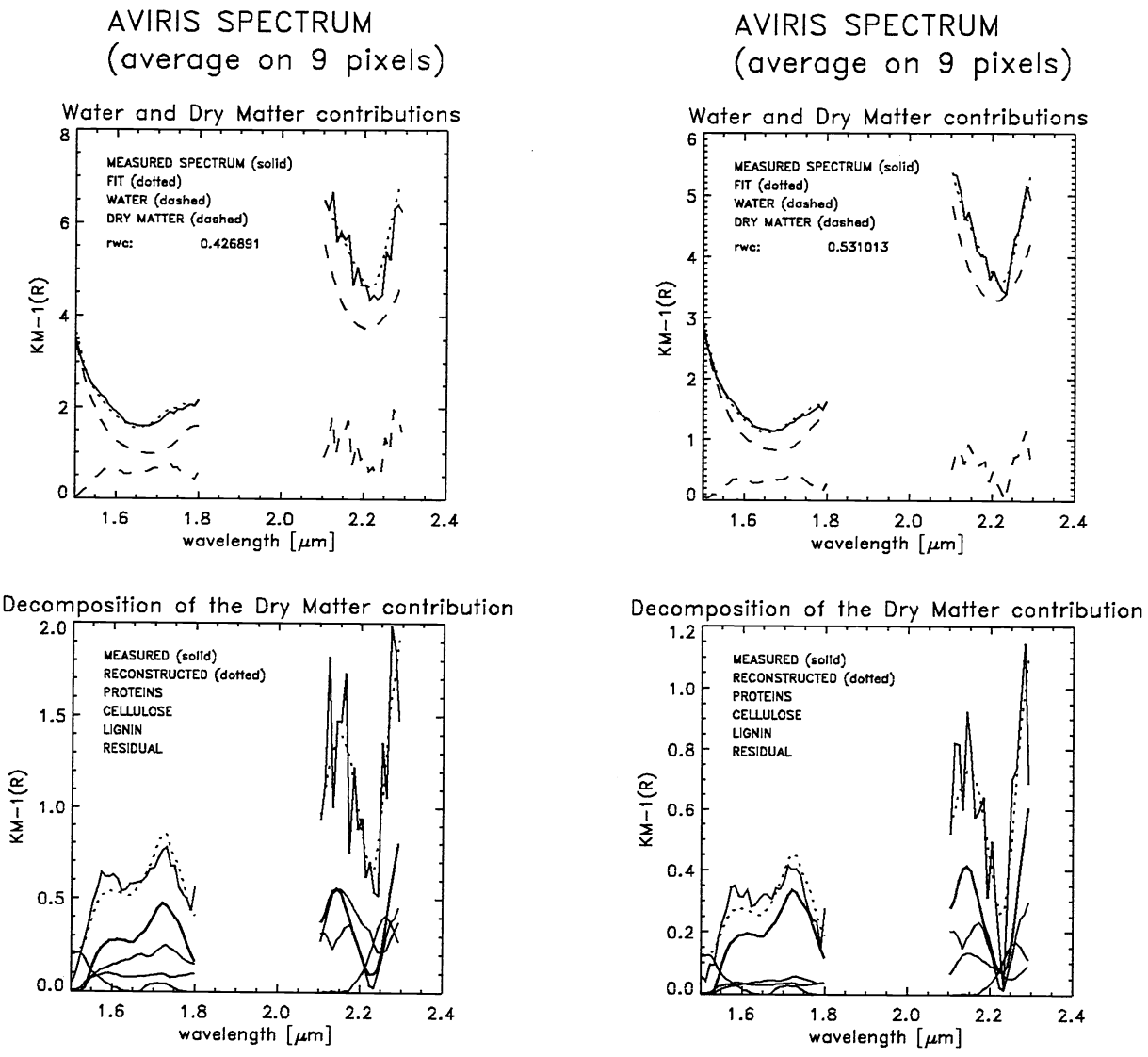
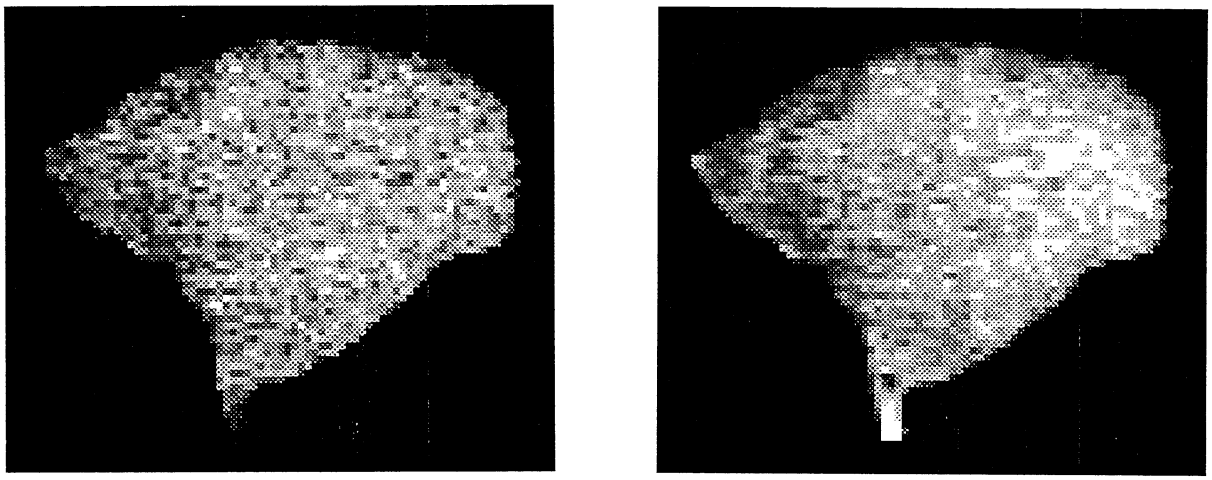


Figure 10. Examples of fitting on AVIRIS spectra from Blackhawk island.



(a) (b)

Figure 11. Blackhawk Island, (a): retrieved relative water content , the range of values is from 0.4 to 0.65
(b): apparent canopy scattering coefficient (refer to text), the range is from 1 to 5 m^{-1}

Figure 10. shows two examples of fitting on AVIRIS spectra. One can notice that the values of $KM^{-1}(R)$ are higher than for the laboratory spectra of stacks, due to the lower canopy reflectance. On the contrary, the spectral shapes are very similar and the values retrieved for the relative water content are very realistic. In Figure 11.a. one can see that this relative water content varies along the east-west direction over the island. The range of values is from 0.4 to 0.6. Figure 11.b. represents the mean (on wavelength) of $1/(q_0 + q_1.\lambda)$, which we have called the "apparent canopy scattering coefficient". It can be considered as a SWIR reflectance brightness corrected for the variation in water content. Further interpretation of these results would need detailed ground information which we do not possess at this moment.

9. CONCLUSION

The investigation of leaf biochemistry by regression analysis provided results which are in good agreement with the literature; the analysis of the optical properties of fresh individual leaves contributes to the originality of this study. We have shown that information on leaf biochemistry was attainable with only a few selected wavelengths. Proteins seem to be the biochemical component that come out the best. However, the calibration equations previously established are not directly usable with remote sensing data.

Using the Kubelka-Munk formula, we have determined empirical "in vivo" absorption coefficients for proteins, lignin and cellulose. Together with the absorption coefficient of water, they allow a very accurate reconstruction of the SWIR spectra of stacked leaves. The inversion of the model so far only allows the retrieval of the relative water content. In this respect, more work is needed both on the estimation of the absorption coefficients and on the inversion procedure.

The inversion of the model on AVIRIS spectra of forests produces apparently meaningful results for the relative water content estimation. These results need to be validated.

10. ACKNOWLEDGMENTS

We wish to thank the NASA HIRIS/Biochemistry team for having provided the 1992 AVIRIS data.

11. REFERENCES

1. Ustin, S.L., Wessman, C.A., Curtiss B., Kasischke, E., Way, J.B., and Vanderbilt, V.C., "Opportunities for using the EOS imaging spectrometers and synthetic aperture radar in ecological models", Ecology 72:1934-1945 , 1991.
2. Peterson, D.L. , "Report on the workshop Remote Sensing of Plant Biochemical Content: Theoretical and Empirical Studies", NASA Ames Research Center, MS 242-4, Moffet Field (CA), 50 pp, 1991.
3. Wessman C.A. , "Estimating canopy biochemistry through imaging spectrometry, in Imaging Spectrometry as a Tool for Environmental Observations", (J. Hill, Ed), Kluwer Academic Publishers (Dordrecht), 1994.
4. Conel, J.E., van den Bosch, J., and Grove, C.I.," Application of a two-stream radiative transfer model for leaf lignin and cellulose concentrations from spectral reflectance measurements", in Proc. 4th Annual JPL Geoscience Workshop, Vol. 1: AVIRIS Workshop (R.O. Green, Ed), Washington DC, 25-29 October 1993, pp. 39-51.
5. Card, D.H., Peterson, D.L., and Matson, P.A.," Prediction of leaf chemistry by the use of visible and near infrared spectroscopy", Remote Sens. Environ. 26:123-147 , 1988.
6. Wessman, C.A., Aber, J.D., Peterson, D.L., and Melillo, J.M. ," Foliar analysis using near infrared spectroscopy", Can. J. For. Res. 18:6-11, 1986.
7. Gastellu-Etchegorry, J.P., Zagolski, F., Mougin, E., Marty, G., and Dubuisson, P., "Forest canopy chemistry with high spectral resolution - An evaluation of the atmospheric influence", in Proc. 6th Int. Symp. Physical Measurements and Signatures in Remote Sensing, Val d'Isère (France), 17-21 January 1994.
8. Curran, P.J., Dungan, J.L., Macler, B.A., Plummer, S.E., and Peterson, D.L., "Reflectance spectroscopy of fresh whole leaves for the estimation of chemical concentration", Remote Sens. Environ. 39:153-166, 1992.
9. Grossman, Y.L., Sanderson, E.W., Ustin, S.L.," Relationships between leaf chemistry and reflectance for plant species from Jasper Ridge Biological Preserve", California, in Proc. 14th Int. Geosci. Remote Sens. Symp. (IGARSS'94), Pasadena (CA), 8-12 August 1994.
10. Lichtenthaler, H.K. , "Chlorophylls and carotenoids: pigments of photosynthetic biomembranes", Methods Enzymol., 148:350-382, 1987.
11. Curran, P.J. , "Remote Sensing of foliar chemistry", Remote Sens. Environ., 30:271-278,1989.
12. Kramer, P.J., Kozlowski T.T.," Physiology of woody plants", Academic Press, 811 pp ,1979.
13. Jacquemoud, S. and Baret, F., "PROSPECT: a model of leaf optical properties spectra", Remote Sens. Environ., vol. 34, pp. 75-91, 1990.
14. Curcio, J.A. and Petty, C.C., "The near infrared absorption spectrum of liquid water", J. Opt. Soc. Am., 41 (5), pp. 302-304, 1951.
15. Wessman,C.A., Aber, J.D., Peterson, D.L., "An evaluation of imaging spectrometry for estimating forest canopy chemistry", Int. J. Remote Sensing, vol. 10, No 8, pp 1293-1316, 1989.
16. Martin, M.E. and Aber, J.D. , "Measurements of canopy chemistry with 1992 AVIRIS data at Blackhawk Island and Harvard Forest",in Proc. 4th Annual JPL Geoscience Workshop, Vol. 1: AVIRIS Workshop (R.O. Green, Ed), Washington DC, 25-29 October 1993, pp. 39-51.